DOCUMENT RESUME

ED 180 431 IR 007 494

AUTHOR Atherton, Pauline
TITLE Online Searching of ERIC: Impact of Free Text or Controlled Vocabulary Searching on the Design of the ERIC Data Base.
INSTITUTION Syracuse Univ., N.Y. ERIC Clearinghouse on Information Resources.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE 79
CONTRACT 400-77-0015
NOTE 235p.; For related document, see IR 007 495; Some tables and illustrations may not reproduce

EDRS PRICE MF01/PC10 Plus Postage.
DESCRIPTORS Annotated Bibliographies; *Comparative Analysis; Data Bases; *Design; Design Needs; *Educational Resources; Information Retrieval; Interviews; *On Line Systems; Research Methodology; Research Projects; *Search Strategies; Tables (Data); Tests
IDENTIFIERS Controlled Vocabulary Searching; *ERIC; Free Text Searching

ABSTRACT

The variety of searchable and printable formats of Educational Resources Information Center (ERIC) data base records has led to different searching methods, which have fostered the confusing state of affairs wherein different results are possible for the same search; this 5-part report describes an investigation of ERIC online use that studied the relationship between free text or controlled vocabulary searching and the design of the ERIC data base. Part I presents the experiences and practices of 45 ERIC searchers who use a number of different retrieval systems. Part II examines the use of free text searching vs. controlled vocabulary searching in over 600 search traces from 18 different ERIC online searchers using four different retrieval systems. Part III uses six search queries from the Online Training and Practice Manual for ERIC Data Base Searchers (ONTAP) to test controlled and free text vocabulary searching strategies. Part IV compares the format of items in the two ERIC printed services (RIE and CIJE) with their online versions in three retrieval systems, makes suggestions for improvements, and reviews previous retrieval effectiveness studies. Part V comprises an annotated bibliography listing 53 citations published from 1970-1978 that deal with online searching using free text or controlled vocabulary. Tables, illustrations, graphs, and figures are used throughout the report. (Author/JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

;free(w)text;free t
FREE(W)TEX
TEXT#FREE:TEXT
ee adj text free

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED180431

ERIC/IR
Special Project Report

ONLINE SEARCHING OF ERIC

Impact of Free Text or Controlled Vocabulary Searching on the Design of the ERIC Data Base

ERIC Clearinghouse on Information Resources
Syracuse University
1979

IR007494

The material in this publication was prepared pursuant to a contract with the National Institute of Education, United States Department of Health, Education, and Welfare. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgment in professional and technical matters. Points of view or opinions do not necessarily represent the official view of the National Institute of Education.

Prepared under Contract Number NIE-400-77-0015

ONLINE SEARCHING OF ERIC:

Impact of Free Text or Controlled Vocabulary Searching on the Design of the ERIC Data Base

An ERIC/IR Special Project

August 1978 to March 1979

Staff

Pauline Atherton, Associate Director, ERIC/IR
and Principal Investigator of Special Project

Karen Markey, Research Associate

Claudia Newton, Research Assistant

Richard Veith, Research Assistant

Michael Rosson, Research Assistant

Marcia Bates, Consultant

Bruce Clark, Project Administrator

Cindy Woolverton, Project Staff

Arlene Sadowsky, Project Staff

Table of Contents

Introduction to Report. . . . . Introduction

Conversations With Online Searchers . . . . . Part I

As They Are--An Analysis of Online Searches Today . . . . . Part II

Online Searching Tests . . . . . Part III

The Redesign of the ERIC Data Base for Online Searching . . . . . Part IV

Online Searching--An Annotated and Analyzed Bibliography . . . . Part V

Consultants for Part III

ERIC Clearinghouse Staff

Sydney Meredith/SO
Pat Spellman/SP
Mike Butler/SP
Norma Howard/PS
Dorothy Bloch/EC
Lynn Smart/EC
Lynn Haupt/HE
Barbara Booth/JC

Places Visited or Called to Speak with Experienced ERIC Online Searchers

Oregon State University, Corvalis
University of Washington
State University of New York at Albany
Health Science Library, Hennipin County Medical Center, Minneapolis
University of Nebraska, Lincoln
Western Michigan University
Boston College
Syracuse University
Kent State University
Northwestern University Library
Lawrence Berkeley Laboratory, University of California
University of Florida, Gainesville
University of Chicago
Pennsylvania State University
CALSPAN Corporation, Buffalo, N.Y.
Educational Testing Service
North Carolina State University
University of Mississippi
University of Missouri, Kansas City
Columbia University
NASIC, Massachusetts Institute of Technology
University of California, Berkeley
University of Utah
University of Kentucky
The Information Store
(FEIS) Florida Educators' Information Service
Stanford University
Deleurs
South Dakota State Library
INFORMS (Iowa Network for Obtaining Resource Materials for Schools)
Georgia Institute of Technology
Tulsa City-County Library
WEECN, (Women's Educational Equity Communications Network)
University of Seattle Health Sciences Library
Rice University
University of Rochester
Carnegie Mellon University
San Diego County, Department of Education
CITE (Coordinating Information for Texas Educators)
ONTERIS
NWREL (Northwest Regional Educational Laboratory)

Thanks for special assistance from:

SMERC
RISE
EPSIS
LRS
SDC
BRS

ERIC/IR
Special Project Report

ONLINE SEARCHING OF ERIC

INTRODUCTION TO REPORT
in
FIVE PARTS

ERIC Clearinghouse on Information Resources
Syracuse University
1979

EXECUTIVE SUMMARY

by Pauline Atherton

The ERIC data base, presently made up of Resources in Education (RIE) and Current Index to Journals in Education (CIJE) data, was first created in 1966 with RIE input only. By 1969, coverage of educational literature published in journals was added by CIJE input. ERIC was conceived as an input/archive system for fugitive literature with two abstracting and indexing publications (RIE and CIJE), and the ERIC Document Reproduction Service (EDRS), the service for reproduction of microfiche and paper copy of available ERIC documents. The ERIC master tape -- the machine-readable records of RIE and CIJE -- has been used for printing cumulative indexes, and for computer-based searching by those persons or groups who subscribe to the ERIC data base or who use the services which emanate from those subscribers. The ERIC data base, in whole or in part (e.g., only RIE or only CIJE), have been available for some time and many computer-based bibliographic retrieval systems, academic information centers, commercial vendors, and state-wide educational research services have used ERIC as their data base rather than produce their own data bases. On occasion, these "tertiary" or "vendor" services have created or added other files, and so their search services have become more diversified than ERIC. WISE, and SMERC, as well as SDC/ORBIT, LRS/DIALOG, and BRS, are cases in point. The recent edition (November, 1978) of the Directory of ERIC Search Services attests to this.

Since Central ERIC has made ERIC tapes available through the ERIC Processing and Reference Facility in Bethesda, Maryland at a very moderate price, there have been limited restrictions on their use. (However, the 1979 CIJE tapes are now copyrighted with limitations.) Plans are now being made for BRS, SDC, and LRS to provide usage reports to Central ERIC. When the ERIC master tape

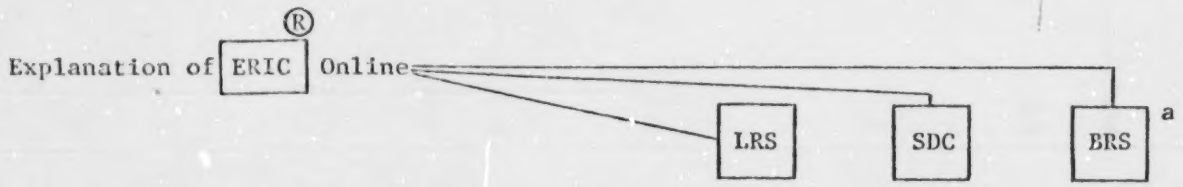
is revised because of vocabulary improvements, addition of new fields, etc., a newsletter (ERIC Interchange) to these subscribers tells them of changes.

Through the years a variety of searchable and printable formats of the ERIC records have appeared and a variety of "Basic Indexes" to ERIC exist in various retrieval systems (see Tables 1 and 2). These tables show how different search results can be obtained if one searches ERIC on SDC, LRS, or BRS. Table 2 is only indicative and should not be construed as a definitive statement on the state(s) of the ERIC data base, but it has caused us to wonder if we should talk of one ERIC data base any longer. Online users of ERIC may indeed be using a different version of ERIC when they switch from retrieval system A to retrieval system B, etc.

More and more online ERIC users are switching back and forth between retrieval systems to search several relevant data bases, or to take advantage of discounts based on volume of use, or to be more efficient/effective because of new retrieval system features. These differences in searching the ERIC data base on different systems becomes very confusing if one makes these switches frequently. Table 3 shows the differences in names of data elements used as search qualifiers by three commercial vendors. The Master Tape data element codes and OCR worksheet names used by ERIC are listed in column 1 for comparison. Rouse and Lannom were among the first to remark about this:

> Given the availability of the same data base on multiple search systems, the searcher's choice of system is likely to be influenced by... important differences in search factors. Search system factors include the default dictionary file, and important search features, such as, truncation, term proximity, and subset searching. Predetermined differences in search system factors were shown to account for differences in results obtained from the data base...
>
> --- S.H. Rouse and L.W. Lannom, On-Line Review, v.1 (1977), p. 117-132. (See Part V of this report for abstract and analysis.)

Table 1. BASIC INDEXES FOR ERIC ONLINE ON LRS, SDC, AND BRS



A. Directly Searchable Fields:

Field Name	LRS	SDC	BRS
Author			AU
Institution			IN
Title	TI	TI	TI
Spon. Agency	SA[b]		SN
Corp. Source	CS[b]		
Descriptor	DE	IT	MJ, MN
Identifier	ID	ST	ID
Note	NT		
Abstract	AB	AB	AB
Command to browse in Basic Index	EXPAND	NEIGHBOR	ROOT

B. Conventions for designating only controlled vocabulary fields in search statement:

	LRS	SDC	BRS
Descriptors:			
Major (full descriptor)	/DF*	*/IT	.MJ.
Minor (or both Major & Minor)	/DF	/IT	.MN.
Single words in descriptor	/DE* or /DE	*/IW or /IW	(Not available)
Identifiers:			
Major (Single words or in full)	/ID*/IF*	*/IT or */IW	
Both Major & Minor	/ID or /IF	/IT, /IW	.ID.

C. Fields not searchable at all

LRS	SDC	BRS
---	availability notes	availability notes grant, report number source

a) In descriptor and identifier fields of BRS record (MJ, MN) only single word descriptors are listed in Basic Index. Single words within multi-word descriptors are not accessed, but plans are underway to change this.

b) LRS plans to delete these fields from Basic Index at time of next update.

Table 2. ANALYSIS OF ERIC DATABASE ONLINE, SEARCHING A SINGLE WORD: READING (as of 3/1/79)

Basic Index	SDC/ORBIT (thru 11/78)	LRS/DIALOG	BRS[a] (thru 11/78)
POSTINGS			
TOTAL:	27,047	27,242	21,959
Field by Field Tally:[b]			
Reading (AU)			13
Reading (TI)	9,293	9,288	9,040
Reading (AB)	??[d]	18,984	18,766
Reading (SN or SA)	---	6[c] (SA)	6 (SN)
Reading (NT)		1,702	
Reading (CS)	---	683[c]	---
Reading.Mn.			1,823
Reading (DE or IW)	20,655 (IW)	20,518 (DE)	
Reading (IT or DF)	2,818 (IT)	2,818 (DF)	
Reading (*IT, *DF, or .MJ) 16,861	966 *IT	996 /DF*	966 /.MJ.
Reading (*IW, or *DE)	15,288	15,248 /DE*	
Reading (ID)	---	1,036	1,024
Reading (ID*)	---	578	
Reading (IF)		0	
SEARCH STRATEGIES TO REDUCE NUMBER RETRIEVED			
1. Limit by Clearinghouse Input:			
Reading & CH = RC	659	660	546
2. Limit by ED or EJ Portion of File:			
Reading & ED	17,910	18,153	15,883
Reading & EJ	9,137	9,089	6,076
3. Limit by Publication Type:			
Reading & DT = B	632	637	477

a) In BRS, plans to allow each part of descriptor and identifier fields to be searchable separately as well as in hyphenated form.

b) See Table 3 describing ERIC Unit Record, BRS/SDC/LRS Search Qualifiers.

c) LRS plans to delete these two fields from Basic Index when they update ERIC files in 1979.

d) To determine tally in abstract field, a time-consuming STRSRCH would have been necessary.

Table 3. ERIC UNIT RECORD - BRS/LRS/SDC Search Qualifiers

ERIC Field Names*	ERIC Codes (RIE/CIJE)	BRS-Search Qualifier	LRS-Search Qualifier	SDC-Search Qualifier
Accession Number	0010	.AN.	---	/AN
Clearinghouse Accession Number	CH 0011	---	---	/CHAN
Clearinghouse Code		.CH.	CH=	/CC
Title*	001A	.TI.	/TI	/TI
Author*	AUTH 001B	.AU.	AU=	/AU
Corporate Source*	INST 001C	.IN.	/CS	/OS
Sponsoring Agency	SPON 0020	.SN.	/SA	/SPO
Descriptor*	DESC 0023	.MJ,MN.	/DE or /DF or /DE* or /DF*	/IT
Identifier*	IDEN 0024	.ID.	/ID or /IF or /ID* or /IF*	/IT
Abstract*	ABST 002C	.AB.	/AB	/BI
Descriptive Note	NOTE 0026	---	/NT	---
Publication Year*	PDAT 0017	.YR.	YR=	FROM__ TO__
Pagination*	PAGE			
Journal Name*	JNL 0032	---	JO=	/JC
Legislative Authority Code		.PA.	AC=	/LPA
Contract/Grant Number	GR 002F CONT 002E	---	CN=	/NU
Bureau/Project Number	BN 0030	---	PN=	/NU
Report Number*	REPNO 002D	---	RN=	/NU
Corporate Source Code	0080	.IN.	SC=	---
Sponsoring Agency Code	0084	.SN.	SC=	---
Document Type*	PUBTYPE 0013	.PT.	DT=	/DT
Group Code	0014(?)	---	GC=	---
EDRS Price	0025	---	---	---
Availability Level	AVAIL 0031	---	---	---
Update	0001	---	UD=	/UP
Issue	002B	---	IS=	/IS

*Several of the fields in the ERIC record are listed in the "Recommended Interchange Specifications for Computer-Readable Bibliographic Data Bases" (April 1978 publication of ASIDIC-EUSIDIC-ICSU-AB-NFAIS). Those found in their list of essential data elements are marked here with an asterisk (*). Their field names vary from ERIC, e.g., Journal is Serial Title, Title is Title of contribution, Publication Date is Date of issue or imprint, etc.

As of January, 1979, ERIC added three new fields (GOV, LANG, and GEO) to the RIE & CIJE Master Tape. These are not in this table because the respective systems have not yet indicated whether or not they will be searchable directly or by prefix codes.

To discover something about the use of ERIC online, we performed a series of studies. First we talked to more than 40 experienced online searchers. Our conversations with them are in Part I of this report.

More than half of these searchers volunteered to show us their activity online. Over 600 search traces were analyzed and are summarized in Part II of this report.

Besides studying their search strategies, we performed tests to investigate differences between controlled vocabulary and free text searching on the Lockheed DIALOG system, the first to have ERIC online. Six queries from the ERIC ONTAP Manual were used. ERIC Clearinghouse staff cooperated in this study (see Part III).

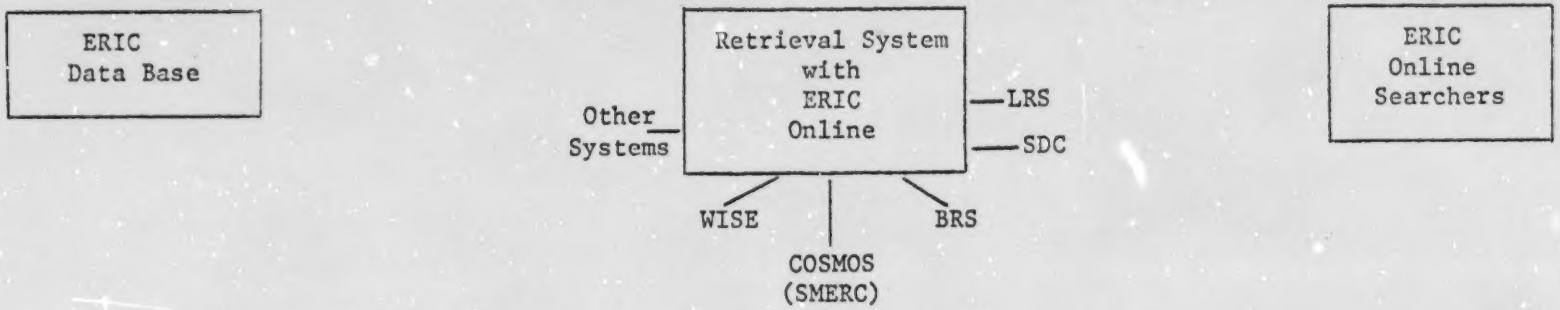
Based on all the above and our own experiences, we have made suggestions for the redesign of the ERIC data base and certain online retrieval system features (see Figure 2 which is a tabular summary and Part IV for details).

Part V contains a bibliography on online searching using free text or controlled vocabulary. It includes just over 50 references of the literature published in the 1970's.

Figure 1 lists the various aspects of the subject of searching ERIC online. In our report we have concentrated on the parts of the ERIC data base and the parts of retrieval systems as well as the needs and consideration of ERIC online searchers.

What follows in this Executive Summary are some extracts to highlight the five parts of the report and a collection of all the suggestions and recommendations for the redesign of the ERIC record, the ERIC online data base and online thesaurus, and the retrieval system features which affect retrieval of ERIC online (see Figure 2).

Figure 1. Aspects of Searching ERIC Online



Parts (Files in Machine-Readable Form)

1) Tapes of RIE, CIJE Records
2) ERIC Thesaurus
3) ERIC Authority File of Identifiers

Back-Up Services:

1) Online ERIC Searcher's Manual and Aids
2) ERIC Master Tape - Record Format & Tape Description
3) ERIC Interchange (Newsletter to ERIC Users)
4) Thesaurus Improvement Project
5) Identifier Clean-Up Project
6) CH Training Workshops & ERIC User Conferences

Features Added to ERIC Data Base:

1) Reformatted master file with data elements of ERIC records changed to fit retrieval system format.
2) Value-added processing of ERIC master file to produce Basic Index (inverted dictionary); decisions about searchable-printable fields. options and variations in print formats; translation of codes.
3) Update or reload ERIC master file from time to time.
4) Use of portions of ERIC Thesaurus tape for access via controlled vocabulary fields, with frequency counts added.
5) Search features which allow access to various elements of ERIC records and allow searcher to set limits on search output.

Back-Up Services:

1) Online ERIC searcher's manual specific to retrieval system
2) System training sessions
3) Update sessions on new system features which affect access to ERIC data base.

Considerations:

1) Amount of training for online searching, ERIC specifically.
2) Experience with ERIC printed products, tools, as well as online.
3) Constraints (lack of uniform training, institutional policies, access, etc.).
4) Relations with End Users which vary search objectives.

Figure 2. Suggestions for Improvements in Searching ERIC Online

(With Responsible Agents Specified)

Responsible Agent (A, B, or C)

A. ERIC System [either Clearinghouse (CH) or ERIC Processing & Reference Facility (F)]	B. Retrieval System	C. ERIC User Group
1) Provide error free records (F).	1) Translate free text searches into controlled vocabulary automatically -- or provide lead-ins between them as in printed ERIC thesaurus.	1) Develop minimum standards for uniform naming of ERIC data elements across systems.
2) Provide ERIC master tape updates if changes are made in thesaurus, identifier field or new data elements are added (F).	2) Provide links between inverted file and controlled vocabulary, similar to rotated descriptor display in printed ERIC thesaurus.	2) Develop requirements for uniform ERIC basic index (inverted dictionary) of single word terms.
3) Provide high quality user aids for ERIC online searchers across all available systems (SDC-LRS-BRS-COSMOS, etc.). (F or CH).	3) Announce planned reloading of files so that ERIC data base suppliers and users can make suggestions.	3) Study types of search objectives to see how they relate to requirements for certain system features and searchable or limitable data elements (e.g., exhaustive search needs fast file, partition capability; high precision needs screening by type of publication, etc.).
4) Provide data elements useful for limiting searches geographically, chronologically, linguistically (F and CH).	4) Confirm to ERIC standard for order of printing data elements in ERIC record and naming of ERIC data elements which are unique. (This pertains to ED and EJ portion of files.)	4) Consider ways to assess and improve searchers' competence
5) Provide "searching thesaurus" which cuts across free text and controlled vocabulary (F and/or CH. Special Project).	5) Provide access to ERIC Basic Vocabulary as recommended by ERIC users. (See A8 & C8).	5) Develop a Review Board for endorsing printed & online user aids (stamp of approval) and repetitive training programs developed by retrieval systems or ERIC system.
6) Consider more precise rules for abstracting to have uniform expressions of action (e.g., implementation of..., effects of..., comparing...). (CH and F).	6) Provide online access to ERIC Thesaurus (all parts) which is at least as good, if not better, than 6B printed version.	6) Work with NFAIS, ASIDIC, and ANSI on cross data base bibliographic standards.
7) Continue studies of free text and controlled vocabulary searching to gather data for vocabulary changes (F and CH).	7) Consider new features to aid online searcher. (See Part IV of full report.)	7) Review differences in access to ERIC online (e.g., pubdate and pubtype) across retrieval systems. Suggest preferred access.
8) Improve ways to limit excessive output (F).		8) Recommend ways to KWIC the ERIC Basic Vocabulary (see A8).
9) Establish combined file of ERIC Descriptors and Identifiers — to be used as basic vocabulary of ERIC online on any system (F).		
10) Run controlled tests to check savings in time and costs using various Search Saves; then develop Public Search Saves for repetitive concepts (where free text and controlled vocabulary are incorporated). (CH).		

Several ERIC Clearinghouses and experienced online searchers of ERIC who use various online systems have created online user guides and search strategy tutorials for both ERIC print and computer searches. We can only list a few of these here.

Table 4. GUIDES/MANUALS FOR THE ONLINE ERIC USER

Bourne, Charles P., et al. Analysis of ERIC Online File Searching. Procedures and Guidelines for Searching. Final Report. (Berkeley, CA: University of California, Institute of Library Research; and Palo Alto, CA: Lockheed Research Laboratories, November 1974). ED 101 757.

Bourne, Charles P. DIALOG Lab Workbook: Training Exercises for the Lockheed DIALOG Information Retrieval Service. (Berkeley, CA: University of California, Institute of Library Research, October 1976). ED 136 751.

Brandhorst, W.T. and Sharon Jewell. Search Strategy Tutorial: Searcher's Kit. (Washington, DC: National Institute of Education, October 1973). ED 082 763.

Clay, Katherine. "Searching ERIC on DIALOG: The Times They are A'Changing," Database 2 (Sep 1979), 46-67.

Markey, Karen and Pauline Atherton. ONTAP: Online Training and Practice Manual for ERIC Data Base Searchers. (Syracuse, NY: ERIC Clearinghouse on Information Resources, Syracuse University, June 1978). ED 160 108.

Mitchell, Kabby, Jr. Preparing Computerized ERIC Search Statements: A Self Instructional Learning Package. (Syracuse, NY: ERIC Clearinghouse on Information Resources, Syracuse University, May 1978). IR-25.

Timbie, Michele and Don H. Coombs. An Interactive Information Retrieval System: Case Studies on the Use of DIALOG to Search the ERIC Document File. (Stanford, CA: Stanford University, ERIC Clearinghouse on Educational Media and Technology, December 1969). ED 034 431.

Yarborough, Judith. How to Prepare for a Computer Search of ERIC: A Non-Technical Approach. September 1975. (ED 110 096, $.83 microfiche, $2.06 photocopy, plus postage, available from EDRS.)

Yarborough, Judith. "A Novice's Guide to ERIC - The Data Base of Education," Online, 1, 3 (July 1977): 24-29.

ERIC/IR
Special Project Report

PART I

CONVERSATIONS WITH ONLINE SEARCHERS

ERIC Clearinghouse on Information Resources
Syracuse University
1979

PART I

CONVERSATIONS WITH ONLINE SEARCHERS

by Karen Markey

Abstract

Several experienced online ERIC searchers were sought out to relate to this research staff some of their experiences and practices. We hoped to learn if there was a difference between habitues of controlled vocabulary or free text searching on either SDC, LRS, or BRS. We thought their clients' objectives or their own training might affect their inclinations toward one or the other search vocabulary. We found some interesting things, but unfortunately, our findings were inconclusive. More careful observations, perhaps on site, should be undertaken. Nevertheless, implications for ERIC are clearer.

Contents

		Page
A.	Introduction	3
B.	Expression of Topics in Search	7
C.	Summary of Observations About Practices Online	8
D.	Search Objectives -- High Precision or High Recall?	9
E.	Issues and Questions	11

Tables and Figures

		Page
Table 1.	Online and Batch Systems Available and Used at 43 Different Search Sites	4
Table 2.	Searchable Facets With No ERIC Descriptors	8
Table 3.	Areas to Investigate in Online Searching	11
Figure 1.	Overlap Among Data Bases Most Frequently Searched by Free Text and Controlled Vocabulary Searchers	5

CONVERSATIONS WITH ONLINE SEARCHERS

by Karen Markey and Claudia Newton

-- We came upon this confession recently in the desk drawer of an erstwhile staff member:

I wasn't intimidated when others accused me of being addicted to searching ERIC online. I eschewed their warnings since I refused to get down until one demonstration of ERIC/ONTAP at the 1978 ASIS Conference in New York. When signing off the terminal, I struggled momentarily. I reached to disconnect the phone, but I couldn't come unhooked! How I feared lest someone trace my trembling hands! Recomposing myself, I limited my concluding remarks to a single statement and ended the interaction with the conferees.

Hastily withdrawing from the ERIC suite, I'd arranged to meet a colleague at a nearby restaurant we'd selected. As soon as I saw her, I blurted out my fear: "I have a terminal illness!"

I recalled those frightening moments, seeking a cure. My buddy explained that I needed to purge ERIC from my thoughts, at least for that evening. She hit the root of my problem: system overflow! We outlined our strategy for the evening. We would not discuss ERIC searching.

Our waiter displayed a certain fondness for me that night, especially when our conversation about ERIC ended. At one point, he expressed in a sort of French accent, "Madam, you have such beautiful eyes. Oh, what prompts Eric's searching elsewhere?"

My buddy honestly tried to keep references to searching offline for the evening. During dessert our curious waiter queried us concerning our stay in New York. Fielding his inquiry, I cited my study of free text at a professional conference. "Ah, free sex," he sighed.

"Yes, at our Clearinghouse we're bound in a probe of online searchers' manipulation of the ERIC data base. We're interested in finding out whether searchers express users' information needs as controlled or free text."

Although he appeared amused, I figured he didn't understand me, so I reformulated my explanation in terms he'd understand.

"We've discussed online searching in phone conversations and face-to-face interviews to online searchers throughout the country and attempted to contact these at institutions serving a diverse clientele; for example, universities, research and development centers, educational information centers, and state education agencies. We requested the procedures they enlist for searching, before and during the online process, with particular emphasis on sources, print and non-print, which they employ to collect searching vocabulary. Involved in isolating search facets or portions of facets which searchers input as free text rather than controlled vocabulary, we queried online searchers about the characteristics of free text search

terms. We encouraged recommendations concerning charges to the ERIC unit record and new retrieval system features in order to enhance ERIC data base searching. For the most part, our discussions with online searchers treated whether they performed searches in controlled or free text."

Puzzled, the waiter made note of his style. "I prefer free sex, but what do you hope to learn from your many contacts?"

"Perhaps that online searchers involved with searching the printed indexes RIE and CIJE are more likely to rely on the controlled vocabulary when retrieving items from the machine-readable data base. Or those involved in performing searches for users in a research environment may employ free text searching since their audience may be dealing with current issues, having terminology not yet integrated into the Thesaurus. We may find out that searchers use free text to satisfy the user's appetite for high recall, i.e., capturing all the information on a given topic. But, to answer a question, there, the controlled vocabulary searching may be sufficient in order to capture a few documents relevant to his/her information needs. We've embraced these aspects as possible outcomes of our free text study. Generally, people have been eager and cooperative participants, particularly those who have submitted their search traces, sometimes called flimsies, for our examination."

Before the waiter withdrew to retrieve our dessert, he noted that he had access to a number of possible participants. My colleague, having assessed our dialogue, reacted with the following response, "ERIC always comes up. You never fail to combine work and play!"

Immediately signing off the topic, I displayed some regret that I'd violated my promise to keep ERIC from entering into the conversation. As we were ready to leave, our waiter inquired concerning my availability.

Recording minor suspicion at my host's statement, I rejected my inclination to build it into a major issue. However, my colleague, as she hurried down the stairs, chided me, "Why didn't you offer him your accession number? You know you've been assigned appropriate descriptors. He might like to come up to see your abstracts. But, of course, your sponsoring agency may get jealous. You'd risk losing your title and identifiers."

Her biting sarcasm stunned me. I staggered alone on deserted streets. Evaluating possible alternatives, I invoked my colleague's strategy to release ERIC from my thoughts hoping that I would be saved temporarily.

---Anon., "Annals of One Online Hooker," (1979).

* * * * * * * * * * *

Any connection between the above and the following report is purely deliberate to add a little humor to a very serious topic.

A. Introduction

Let me discuss in detail the highlights which proceeded my experience at the New York ASIS Conference. Tracking down online searchers at the annual meeting was valuable, but it pointed out that efforts to contact a geographically dispersed population affiliated with a variety of organizations would have to be accomplished by telephone rather than personal interviews. Yet, one-quarter of the 42 online searchers contacted by the project staff was questioned in face-to-face interviews.

Our* questioning of online searchers is termed interviewing in this report. Interviewing is used loosely to describe the means we employed to find out how searchers manipulated the ERIC data base online. It must be emphasized that our intent was not to adhere to a prescribed set of questions; rather, to identify and pursue fruitful areas of inquiry. Summarized in this account are the responses of 29 searchers from postsecondary educational institutions: ten from state and county level education agencies, three from private enterprises, two from research and development centers and one searcher from an educational information center. Our objective throughout the questioning process was to accommodate for and encourage respondents who deviated from a fixed interviewing plan. We were primarily concerned with gathering suggestions for ERIC unit record format, retrieval system features, and implications for subsequent investigations rather than with answers to every question..

Only seven searchers had access to ERIC through a single retrieval system. DIALOG and ORBIT were reported as the most frequent combination with sixteen users. Fifteen respondents had access to retrieval systems besides DIALOG or ORBIT or BRS. A summary of systems available to those contacted is given below.

TABLE 1. Online and Batch Systems Available and Used at 43 Different Search Sites.

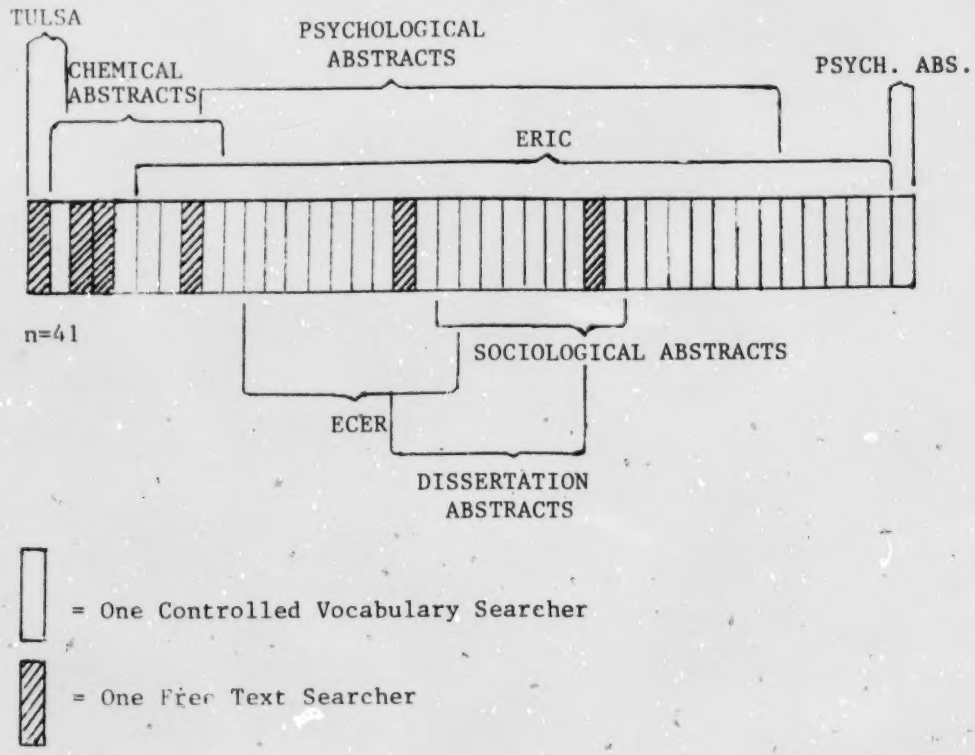
System	Number (n=45)
Lockheed Information Systems	36
System Development Corporation	31
Bibliographic Retrieval Systems	23
New York Times	7
National Library of Medicine	6
Batelle	2
Department of Energy	2
Technotec	2
Othe.*	8

Eight searchers did not include ERIC among the list of data bases they searched frequently, citing data bases in the fields of medicine, chemistry, business, and geology. Three of the eight reported that they performed primarily free text searching of the ERIC data base. Since a total of six searchers responded that they employed free text for searching the ERIC data base online, there are three free text searchers among the remaining 35 respondents who searched ERIC frequently. Twenty-seven of these 35 also searched Psychological Abstracts. Exceptional Child, NTIS, Dissertation Abstracts, and Sociological Abstracts were also mentioned by ERIC searchers as files they accessed often.

Let's take a closer look at these six online searchers who have described themselves as free text searchers, in order to determine what aspects distinguish them from searchers performing descriptor searches. One of the six did not own the Thesaurus of ERIC Descriptors. Although she checked RIE or CIJE in her search routine or called up DIALOG's online ERIC Thesaurus, she input search terms in free text. One searcher who performed descriptor-based searches also did not own a Thesaurus. But she used the online thesaurus in order to construct formulations in controlled vocabulary.

*Systems reported once are as follows: COSMOS, DOW JONES, Georgia Information Dissemination Center, PIRETS, RECON IV, TRIS, Univ. of Kentucky batch processor.

FIGURE 1. Overlap Among Data Bases Most Frequently Searched by Free Text and Controlled Vocabulary Searchers*



All searchers consulted some printed or non-printed sources of controlled vocabulary to gather searching vocabulary. Two of the four free text searchers responded when asked whether they used CIJE and/or RIE that they check these sources for index terms, especially for known relevant citations, or when the user was uncertain about the term usage of his area of interest. All but five descriptor searchers used the printed indexes during presearch preparation in the same way as free text searchers: to check for term usage, or find relevant citations in order to scan their indexing records or titles for suitable terms. Three searchers consulted the printed indexes in order to assess whether to per-

*Two of the 45 searchers contacted were not questioned concerning ERIC searching and are not included in the figure above. The remaining two searchers use Medline; they are not frequent ERIC searchers.

form an online search on the topic.

Except for the two searchers who did not own an ERIC Thesaurus, all searchers used the printed ERIC Thesaurus prior to signing on the terminal. Comments from free text searchers varied: "to formulate strategy", "'dangerous' not to use it", "to set up search and contingency situations beforehand", or "not my field, so I feel comfortable getting all alternatives".

Descriptor searchers responded with similar remarks which have been classified: to find postings information, select correct terms and spelling, scan RT's and BT's, to formulate search strategy, and to save time. What must be stressed is that all searchers, whether they input search terms in free text or controlled vocabulary, enlisted the printed or online ERIC Thesaurus before or during their online search.

However, if no descriptors were available for representing a search concept, did free text searchers select printed sources different from those employed by descriptor-based searchers in order to cull searching vocabulary? The Thesaurus of Psychological Terms, Roget's Thesaurus, and dictionaries were cited by free text searchers as well as those who performed descriptor searches. They offered additional sources: education dictionaries, Education Index, DIALIST, ABI/INFORM's Thesaurus, Management Content's Thesaurus, in-house morgue of topics, or a bibliography in the field. Both types of searchers relied on the requestor's account of his question, or consulted an expert in the field. Searchers enlist the online data base in order to cull search terms. Based on the requestor's suggestions, the searcher may call up a known relevant citation, scan the indexing record, title, or subject fields in order to check for appropriate terminology. This approach to online searching, called Citation Pearl Growing, makes the most use of the interactive capabilities of the online search system.

Some free text searchers will start with one known item; others will start with the highest posted concept to build the largest set of items which they can then "fractionate" by other concepts. All these strategies are described in the ERIC ONTAP Manual by Markey and Atherton (ED 160 108).

B. Expression of Topics in Search

Those we contacted agreed that not all searchable topics in ERIC were capable of being expressed in ERIC's controlled vocabulary. Before entering a term as free text, descriptor searchers were more likely to select descriptors similar in meaning to the concept at hand, scan and select from the output on-line for possible relevant hits, then input controlled vocabulary from possibly relevant citations. Another strategy involved checking for identifiers on the topic of interest, or finally, inputting the topic as free test. With one exception, all descriptor searchers selected free text as their preferred strategy for treating new topics in Education which were not integrated into the controlled vocabulary. The single disagreeing searcher sought descriptors assigned to relevant articles which he found, for example, by an author search. As the only ERIC searcher using SDC exclusively whom we contacted, he was more inclined than other respondents to exhaust every possible strategy for identifying appropriate descriptors before searching the basic index.

It is noteworthy that all descriptor searchers chose to employ free text in instances when no controlled vocabulary was suitable for a given concept. Throughout the interviews, we sought to isolate and classify facets for which no controlled vocabulary existed. Online searchers offered the following suggestions:

TABLE 2. Searchable Facets With No ERIC Descriptors

1. Geographic areas -- This is not limited to retrieving items in states or cities, but regions, e.g., Pacific Northwest, Upstate New York.

2. Recent topics in Education -- for example, generalization of alternatives, zero based budgeting

3. Specific named objects -- for example, test names, diseases

4. Value judgments -- for example, pros and cons, or advantages and disadvantages of a given treatment

5. Action statements -- What effect one thing has on another, or the relationships between or among given items, e.g., implementation of . . . , problems associated with . . . , effects of . . . , comparing . . . Descriptors for evaluation, and descriptors and publication type for research have been mentioned by respondents as incapable of specifying relationships or action.

These represent areas named by respondents during interviews. However, in Part IV we provide additional insight based on a random sample of search traces in which free text is used. Here, free text search terms are examined in an effort to classify them into the above categories, or amend the list of facets.

C. Summary of Observations About Practices Online

Reviewing the foregoing discussion points out that both free text and controlled vocabulary searchers were not restricted to a particular organization, that they used both online and printed thesauri, and the same alternate sources for gathering search terms not in the controlled vocabulary. All searchers employed free text when no controlled vocabulary existed, particularly in the case of new topics in Education.

Responses by free text searchers concerning whether user objective, e.g., recall, precision, immediacy, etc., influenced their decision to express the

search in controlled vocabulary or free text, centered on the ability of controlled vocabulary searching to deliver high precision output. It is noteworthy that free text searchers were aware of high precision output as a result of descriptor-based searching. However, the attitude of one free text searcher was reflected in her statement, "Most people seem to want alot, 'the more the merrier' in social science, it seems. Never had anyone who wanted a precise search."

When questioned concerning problems associated with free text, four out of five free text searchers replied that high volume of postings could usually be anticipated and avoided or limited. They answered in a similar way concerning the precision of free text searches, that false drops were very much likely and expected to occur.

D. Search Objectives -- High Precision or High Recall?

When interviewing two searchers simultaneously, one used a combination of free text and controlled vocabulary in an effort to retrieve as much as possible on a given topic. Her associate, however, formulated high precision strategies, believing that the end user would be overwhelmed by a large output. Thus, at a single institution, one searcher interpreted the user's objective as high recall, the other as high precision. On the other hand, when interviewing groups of searchers at two state education agencies, it became clear that the searchers' objectives were to create a high precision search because the output would be screened for relevance and appropriate items delivered to the end user. One searcher noted that controlled vocabulary was used in order "to answer a question", and in most instances it provided a manageable number of relevant citations from which the "answer" was construed. Another searcher cited that controlled vocabulary, especially when it "mapped" onto a topic,

provided an easy way to get a "reasonable" amount of material.

Controlled vocabulary searchers seemed more likely to compose high recall searches in a combination of free text and controlled vocabulary. One interesting comment by a descriptor searcher, that "sometimes free text is even more precise than the descriptors", reveals attitudes expressed by nearly one-third of descriptor searchers. In this instance, free text is used to indicate a concept not in the controlled vocabulary, particularly a new topic in Education or a named object. Rather than using controlled vocabulary, descriptor searchers sometimes noted that they obtained a few relevant citations by searching the title or identifier fields. For the most part, there was general agreement among all searchers regardless of their orientation that high precision resulted from descriptor-based searching. Likewise, free text increased recall.

No characteristic differences between types of searchers have been pointed out; rather, similarities have emerged from this discussion. For example, all consult the ERIC Thesaurus, and searchers from both groups check dictionaries in Education, Thesaurus of Psychological Terms, CIJE, and RIE in order to gather search terms. Free text searchers were not limited to a single organization, and have access, like controlled vocabulary searchers, to one or more commercial systems, or additional retrieval systems on which ERIC was unavailable. Investigating the user's objective brought out some interesting points; for example, more than one controlled vocabulary searcher asserted that controlled vocabulary searching, and to a greater degree, controlled vocabulary and free text searching, produced too much material so their task was to limit the output. Curtailing the output was usually accomplished by applying a limit to major descriptors. It may be concluded then that controlled vocabulary searching yielded a suitable number of citations to satisfy the user's needs. However, controlled vocabulary searchers, cognizant that free text searching was necessary to provide access to fields besides those having controlled vocabulary,

used this capability to retrieve citations treating concepts not available in the controlled vocabulary.

For the most part, controlled vocabulary searchers were concerned about overwhelming users with voluminous output. On the other hand, free text searchers anticipated a high volume of postings and false drops, and accounted for this in their strategy. Like controlled vocabulary searchers, they sought to achieve their user's objective, but were more likely to employ free text rather than controlled vocabulary. As one searcher stated, "The more, the merrier, in social science."

E. Issues and Questions

The issues explored by the interviews served to raise more questions. Subsequent studies might investigate the following aspects in an attempt to find more contrasts between free text and controlled vocabulary searchers.

Table 3. Areas to Investigate in Online Searching

1. User's objective(s) (high recall or high precision) -- The Free Text study treated this issue, but future investigations may take into account objectives and some other factors, for example, cost and user's purpose.

2. Searcher's perceived notion of the user's objective -- Has the searcher found out the user's objective, or enacted his own perception of what the objective ought to be?

3. Type of organization -- policy of the institution concerning computer-based searching, newness of searching services.

   a. Clientele -- This may affect subject areas which are searched. A research organization may be involved with recent topics in the field.

   b. Client's purpose -- Here, the user's objective is revealed, or reckoned by the searcher.

   c. Document screening and delivery -- Like the state education agencies discussed above, searchers who have to read the output, assess relevance, and provide hardcopy, may be interested in high precision searches in order to curtail non-relevant citations.

   d. Cost -- If the connect time and offline print charges are passed on to the end user, does it have an effect on the searcher's decision to choose controlled vocabulary for high precision searching?

4. Search as an assessment of the searcher's competence -- If a searcher produces a high recall search having much non-relevant material, does this reflect his incompetence as a searcher?

5. Frequency of searching the ERIC data base online -- Does a searcher employ searching techniques like free text input when he/she is accustomed to using it when searching other files? A searcher who is unfamiliar with the controlled vocabulary in Education may or may not use it when searching.

6. Training for computer-based reference services -- What was the state-of-the-art when the searcher was trained?

   a. Features available on the training system -- If a searcher learned techniques before the introduction of free text searching, is he/she more likely to use controlled vocabulary?

   b. Policy of institution providing the training -- This involves both the institution and the teacher.

7. Volume of output -- Do searchers anticipate a sort of "ideal" final set size before they being a search or for all searches?

ERIC/IR
Special Project Report

ONLINE SEARCHING OF ERIC

PART II

AS THEY ARE:
AN ANALYSIS OF ONLINE SEARCHES TODAY

ERIC Clearinghouse on Information Resources
Syracuse University
1979

PART II

AS THEY ARE:
AN ANALYSIS OF ONLINE SEARCHES TODAY

by Karen Markey and Claudia Newton

Abstract

Over 600 search traces from 18 different ERIC online searchers using four different retrieval systems are examined in depth to determine the use of free text searching vs. controlled vocabulary searching. Terms used for searching and for output are differentiated. The findings show that controlled vocabulary terms are used in 68% of the ERIC searches. Free text searching is lower if only terms which resulted in output are considered. Analysis of average number of terms in search/output for three types of searches is shown. Comments are made about the use of features of the three retrieval systems used by different types of searchers. Those features affecting free text searching are highlighted. The vocabulary in 165 free text search statements were analyzed to determine why ERIC descriptors were incapable of representing these concepts. One out of eight statements were not able to be represented in controlled vocabulary, but careful analysis showed that more than half of these could be re-expressed in controlled vocabulary.

Contents

		Page
	INTRODUCTION	1
	A. What is a Search? An Output?	2
	B. What is a Free Text Term? A Controlled Vocabulary Term?	4
I.	FINDINGS ABOUT SEARCH TRACES	6
	Summary and Recommendations for Further Research	13
II.	RETRIEVAL SYSTEM FEATURES	14
	A. Word Adjacency	15
	B. Word Proximity/Sentence Searching	16
	C. Field Searching	16
	D. Single Free Text Terms	16
	E. Truncation	17
	F. Controlled Vocabulary	17
	G. Positional Operators and Truncation Searching	18
	H. Multiple Positional Operators	18
	I. Online Dictionary	18
	J. Online Thesaurus	19
	K. Citation Searching	19
	L. Conclusions	20
III.	VOCABULARY STUDIES OF FREE TEXT SEARCHES	21
	Summary of Findings About Vocabulary Usage	28

Contents (Con't.)

Tables

		Page
Table 1.	Retrieval Systems Used by Online Searchers.	1
Table 2.	Contents of Basic Index for ERIC Data Base in Three Retrieval Systems	4
Table 3.	Summary of Statistics about Search Traces	7
Table 4.	Search Features of Retrieval Systems Reviewed by Project Staff	15
Table 5.	Terms with Descriptor and Free Text Postings.	22
Table 6.	Lead-in Term and Descriptor Postings for Ten Terms.	23
Table 7.	Free Text Statements with Variant Descriptors	24
Table 8.	Free Text Statement in Search with Possible Descriptors	25
Table 9.	Free Text Statements with Postings in Identifier Field and other Free Text Fields.	26

Graphs

		Page
Graph 1.	Use of Free Text by Online Searchers.	8
Graph 2.	Percentage of Free Text Within Search/Output.	9
Graph 3.	Percentage of Free Text in all Searches Comparing Free Text	10
Graph 4.	Number of Search Terms Used by Three Types of Searchers	11
Graph 5.	Number of Free Text Search/Output Terms	12

Figures

		Page
Figure 1.	ERIC Online Searches--Percentage of Input as free or controlled vocabulary	31

AS THEY ARE:
AN ANALYSIS OF ONLINE SEARCHES TODAY

INTRODUCTION

When talking to online searchers, we requested that they submit to us a sample of their traces from ERIC searches.* We emphasized that we wanted a sample, not unusual, outlandish strategies or topics, just a "typical" batch of histories from a week's requests. We stressed that we were not interested in evaluating their proficiency as an online searcher, in subject or content of the search, or the client's identity or purpose. There was only casual mention concerning how we intended to use the traces. In fact, searchers inquired neither about our purpose nor procedures for analyzing the histories.

Response was overwhelming. Over 20 searchers submitted traces. This report contains results on traces obtained from 18 online searchers. One collection of traces was eliminated because the searcher cut off portions of the trace necessary for our analysis (see What is a Search? An Output? below). Other collections were disregarded since they were received too late.

The following table summarizes retrieval systems used by searchers and is classified according to the searcher's affiliation.

Table 1. Retrieval Systems Used by Online Searchers

	Contacted	# Who Submitted Traces	Retrieval Systems LRS	BRS	SDC	COSMOS
Universities	28	10	7	3	-	-
State Education Agencies**	9	5	3	3	-	1
Information Brokers	3	0	-	-	-	-
Educational Info. Centers	1	1	1	-	-	-
Research & Devel. Centers	2	2	2	-	-	-
TOTAL	43	18	13	6	0	1

*See separate section of this Project Report, "Conversations with Online Searchers".

**Two SEA's submitted traces from both LRS and BRS.

A. WHAT IS A SEARCH? AN OUTPUT?

In order to analyze the traces, we sought a definition of search. Bourne et al. (1974)*, in their examination of online searches, found it necessary to distinguish between search and question when they analyzed a tape-log for 15 days searching at nine terminals produced by LRS. "Search" was initiated by the BEGIN command or a system-generated message of I/O SUBTASK TERMINATED. A "question" was considered complete in the above mentioned instances, or when the END command was input. Thus, one search may be comprised of one or several questions.

In our examination of search histories, we make a distinction between search and output. For the most part, our focus is on searching vocabulary, i.e., have search terms been entered as free text or controlled vocabulary? Our initial efforts centered on counting and describing all search terms input by the searcher. However, it became clear that the total number of terms ought to be distinguished from those which formed the output delivered to the end user. In this way, we are accounting for the interaction between system and searcher --- The searcher's strategy develops prior to signing on, and continues to

Formulation		Explanation
? S GIFTED 1	114 GIFTED	The Searcher has only chosen four free text terms for this search : GIFTED, TALENT? IDENTIF?, and ARTIST?.
? S TALENT? 2	113 TALENT?	
? C 1 OR 2 3	195 1 OR 2	
? SIDENTIF? 4	2262 IDENTIF?	
? C 3 AND 4 5	45 3 AND 4	In search statement 7, the searcher retrieved one citation when the term ARTIST? was combined with postings from GIFTED, TALENT?, and IDENTIF?.
? S ARTIST? 6	156 ARTIST?	
? C 5 AND 6 7	1 5 AND 6	
? PRINT 5/5/45		The output ordered comprised retrieval from the first three terms only.

*Bourne, Charles P. et al. Analysis of ERIC Online File Searching. Procedures and Guidelines for Searching. Final Report. Berkeley, CA: University of California, Institute of Library Research, and Palo Alto, CA: Lockheed Research Laboratories, November 1974. (ED 101 757)

evolve as the vocabulary is input, postings are evaluated, citations are scanned for relevance, and so on, until the decisive point where the searcher orders offline or online prints, or terminates the search. An example of search and output is presented above with explanations.

Over 600 search traces were examined by our project staff. Counts of and descriptive information for search terms were recorded on search trace tally sheets for the following:

(A) SEARCH

1. total number of terms comprising the search
2. number of major descriptors
3. number of minor descriptors
4. number of free text terms
5. fields specified for (A)4

(B)
1. total number of terms comprising the output
2. number of major descriptors
3. number of minor descriptors
4. number of free text terms
5. fields specified for (B)4

In-depth analysis of all free text searches was done by monitoring usage of retrieval system features associated with free text searching.

B. WHAT IS A FREE TEXT TERM? A CONTROLLED VOCABULARY TERM?

In order to define a free text term we had to take into account how every retrieval system processes the ERIC unit record. From the table below, it is clear that each basic index in the three major retrieval systems is comprised of a different set of subject-conveying fields. Only the title, abstract, and identifier fields are contained in the basic indexes of all three systems. Other fields, descriptor, note, sponsoring agency, corporate source, and author are treated differently.

Table 2. Contents of Basic Index for ERIC Data Base in Three Retrieval Systems

Retrieval System+			Field
ORBIT/SDC	DIALOG/LRS	BRS	
/AU	AU=	AU	Author
/OS	CS	IN	Corporate Source
/SPO	SA	SN	Sponsoring Agency
TI	TI	TI	Title
AB (String-search)	AB	AB	Abstract
ST (String-search)	ID,IF,ID*,IF*	ID	Identifiers
IT,IW,*/IT,*/IW	DE,DF,DE*,DF*	MJ,MN	Descriptors
NO (String-search)	NT	Not Search-able	Note

CONTROLLED VOCABULARY

FREE TEXT

+Basic Index of each system is the boxed portion.

Single words in all these fields are listed plus descriptors in full.

Controlled vocabulary terms are defined as those assigned to express the item's intellectual content. Only terms from the major and minor descriptor, and identifier fields make up the controlled vocabulary. Free text terms are derived from subject-conveying fields of the ERIC unit records: title, sponsoring agency, corporate source, note, and abstract. Terms from the descriptor and identifier fields are counted as free text when they are retrieved simultaneously with subject-conveying fields.

DIALOG's basic index takes into account all subject-conveying fields. BRS' basic index also contains the author field which would be regarded as a coded search in both ORBIT and DIALOG. For example, what may obviously appear to be an author search in BRS (e.g., LIPPINCOTT ADJ PAUL) would be counted as free text when the statement was not terminated by field delimiters (e.g., LIPPINCOTT ADJ PAUL.AU.). In this instance, there is no way to determine whether citations were retrieved only from matches on the author field or other subject-conveying fields since the author field is processed into the basic index in BRS. Likewise, coded searching in ORBIT in the corporate source or sponsoring agency fields (e.g., READING/SPO) would be counted as free text since the searcher has designated retrieval from a subject-conveying field, which in this case, is not processed into ORBIT's basic index. It is important to note that our definition of search terms centered on subject-conveying fields rather than what each retrieval system designated as contents of the basic index.

We were interested in finding out what percentage of ERIC searches employ free text terms and comparing this with what searchers described about their representation of search terms as input to the retrieval system. It was necessary to establish some guidelines concerning how to count terms. The following examples serve to illustrate typical cases:

	Comments
1. TALENT IDENTIFICATION (or) TALENT-IDENTIFICATION	1. one controlled vocabulary term
2. EDUCATIONAL-0$	2. one controlled vocabulary term, but it retrieves all descriptors in the rotated display from Educational Objectives to Educational Trends
3. EDUCATIONAL/DE	3. one controlled vocabulary term, but it retrieves 68 descriptors which contain the term EDUCATIONAL
4. FOUR ADJ H.ID. (or) FOUR(W)H/ID (or) FOUR H	4. One controlled vocabulary term
5. FOUR ADJ H (or) FOUR(W)H	5. two free text terms
6. INDO(W)EUROPEAN(F)LANGUAGES. (or) INDO ADJ EUROPEAN WITH LANGUAGE$1 (or) INDO AND EUROPEAN AND LANGUAGE#	6. three free text terms, although statements in SDC and LRS would retrieve the descriptor INDO EUROPEAN LANGUAGES
7. LANGUAGE$ (or) LANGUAGE? (or) ALL LANGUAGE#	7. one free text term, but 107 LANGUAGE(S) descriptors are retrieved in SDC or LRS
8. ZERO ADJ BASED.ID,TI. (or) STRS(TI,ST):ZERO#BASED: (or) ZERO(W)BASED/ID,TI	8. two free text terms

I. FINDINGS ABOUT SEARCH TRACES

Six hundred fifty seven traces were examined. The sets of search traces submitted by 18 searchers ranged from three to 402 traces. The figures below report sums of percentages divided by the number of search history sets so that no set outweighs others.

Controlled vocabulary terms comprised nearly two-thirds of an ERIC search. The percentage of free text terms among the 18 sets varied from 4% to 100%. An output, that portion of the search which makes up the printed citations, was comprised of 68% controlled vocabulary. Only 3 of 18 searchers expressed the output in a greater percentage of free text terms than the search.

Table 3. Summary of Statistics about Search Traces

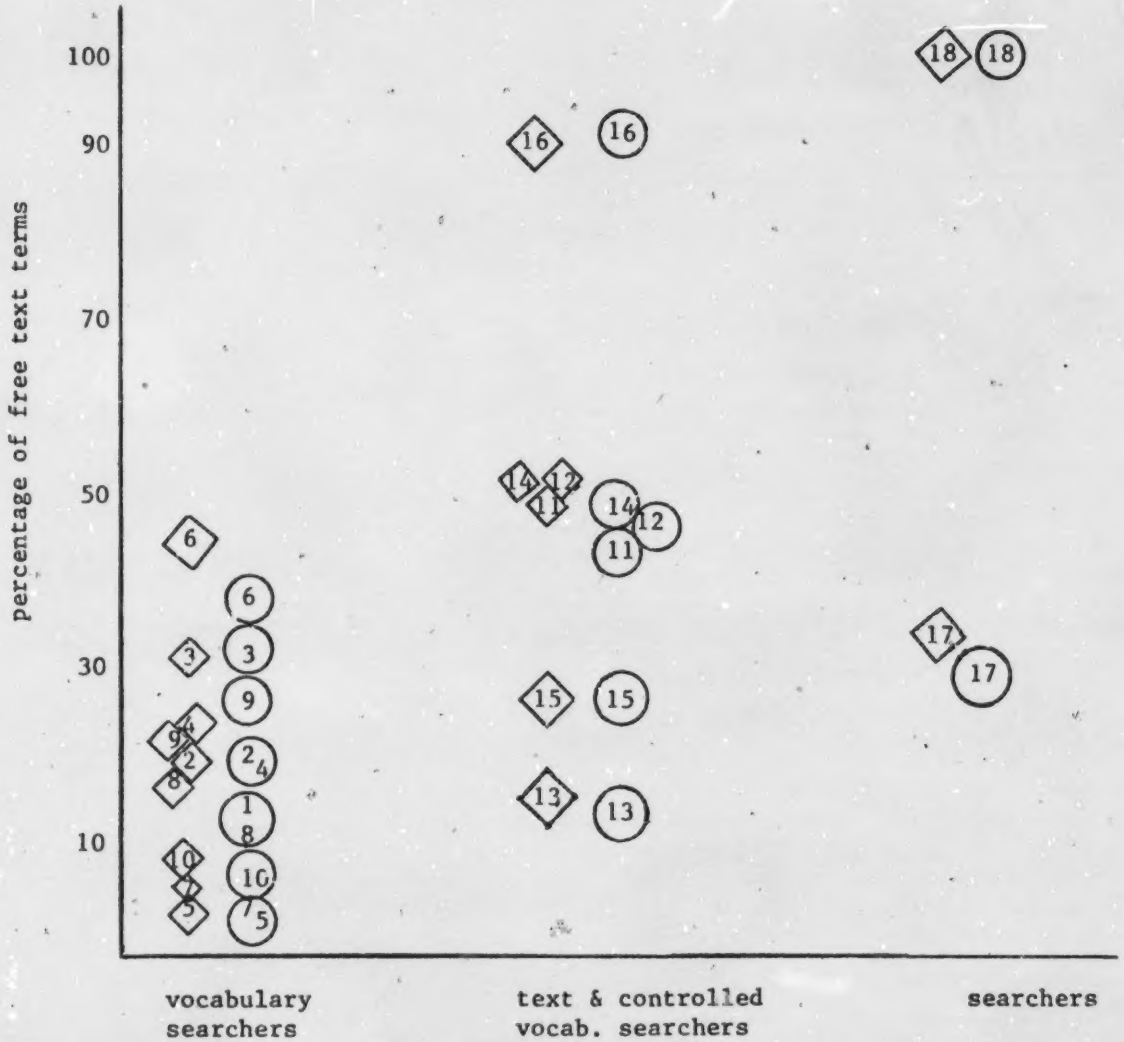
	Search	Output	Range per Set of Search Traces
1. number of searches	657	643	3 to 402
2. % of free text terms per search/(output)	35	32	4 to 100
3. % of the number of searches/ (output) using free text	64	58	18(16) to 100
4. average number of free text terms per search/(output)	6.8	5.9	.3 to 32.6
5. average number of terms per search/(output)	16.5	14.7	6.93 (5.9) to 32.7 (32.6)
6. average number of free text terms in searches/(outputs) with any free text	8.9	8.2	1 to 32.6
7. average number of terms in searches/(outputs) with any free text	17.3	15.6	6.4 (5) to 37.1 (32.6)
8. averaged % of the average number of free text terms per search/(output) with any free text to average number of free text terms	48.6	49.2	10 to 100

We might speculate that volume of postings or false drops could be associated with this difference, that the searcher chose to express output in more controlled vocabulary after reviewing online citations and assessing their relevance. In searches containing any free text, the output was likely to be comprised of nearly half free text terms.

During our interviews of online searchers, we asked them whether they performed primarily descriptor searches. Based on their responses, searchers were separated into three groups: controlled, combination, and free text searchers. The following graphs illustrate how these three types of searchers input vocabulary to the retrieval system. The numbers found on the graph refer to our numbering of the sets of traces received from 18 online searchers.

GRAPH 1 -- USE OF FREE TEXT BY ONLINE SEARCHERS

Searchers have been organized into three groups. When the percentage of free text terms per search/output is designated for each searcher, an increase in the percentage of free text terms is evident, especially for combination and free text searchers. Note that the outputs of seven of the ten controlled vocabulary searchers contain less than 20% free text. Only one output of the six combination searchers was less than 20% free text. A slight decrease in percentage of free text in the output occurs among the three types of searchers.



THREE TYPES OF ERIC ONLINE SEARCHERS

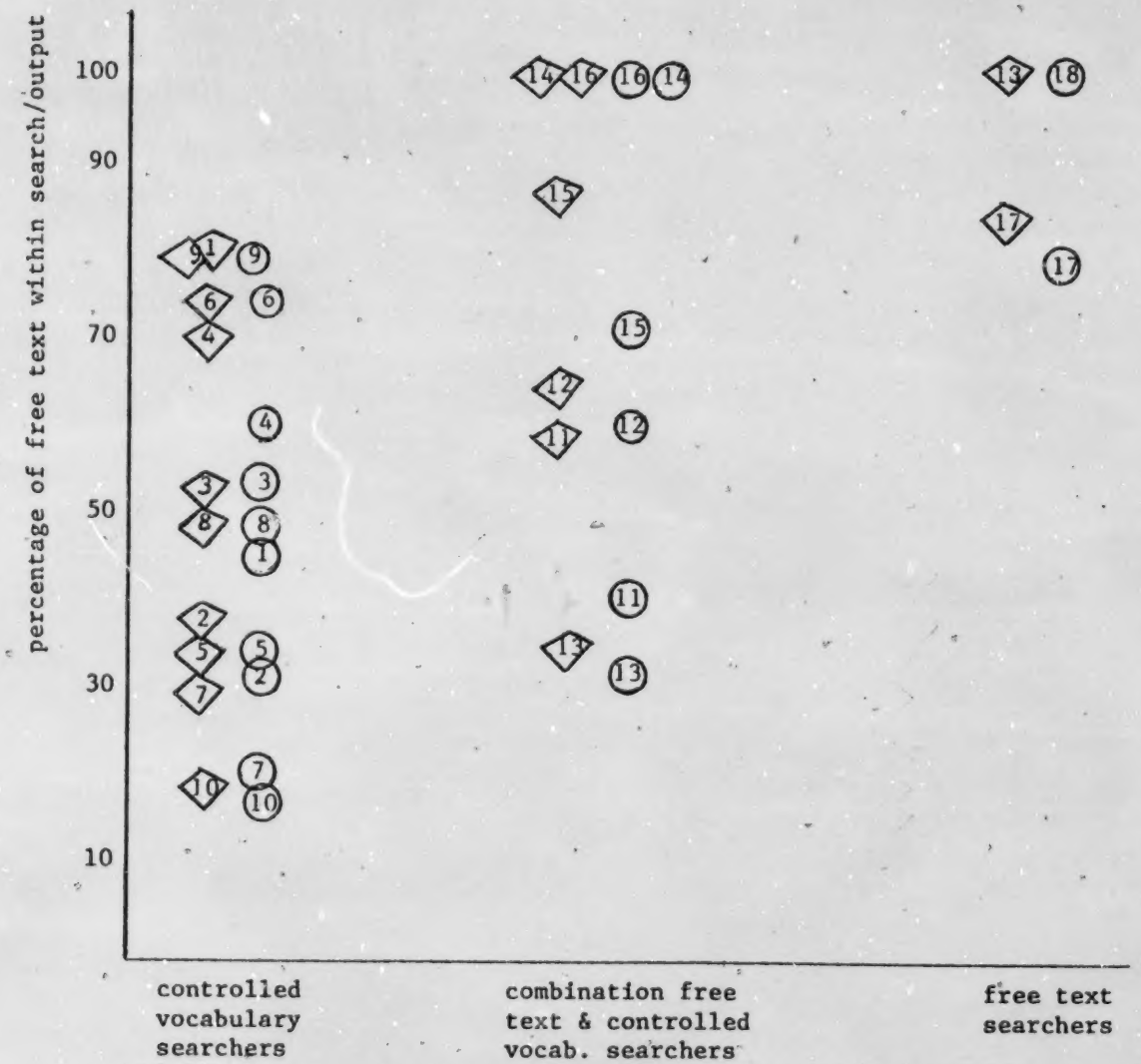
◇n = search

(n) = output

The numbers found on the graph refer to our numbering of the sets of traces received from 18 online searchers.

GRAPH 2 -- PERCENTAGE OF FREE TEXT WITHIN SEARCH/OUTPUT

This shows the percentage of free text per search/output in those searches which have any free text. Instead of a decrease in percentage of free text comprising output, it is clear that output contains a slightly greater percentage of free text terms than the search.



THREE TYPES OF ERIC ONLINE SEARCHERS

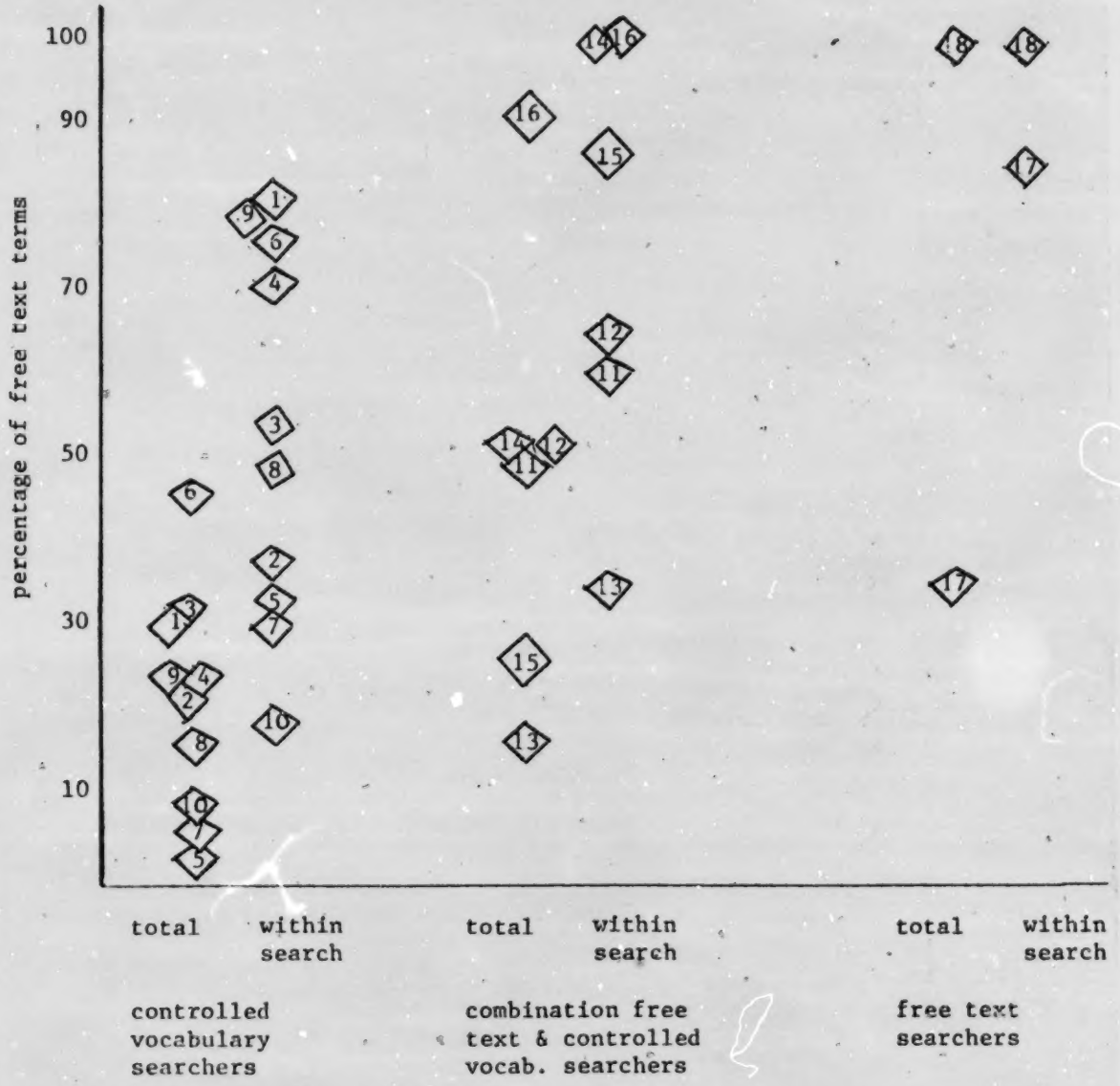
search (n)

output (n)

The numbers found on the graph refer to our numbering of the sets of traces received from 18 online searchers.

GRAPH 3 -- PERCENTAGE OF FREE TEXT IN ALL SEARCHES COMPARING FREE TEXT WITHIN SEARCHES

Overlapping Graphs 1 and 2 demonstrates that all three groups of searchers are likely to express a greater percentage of the search in free text when any free text is contained within the search.

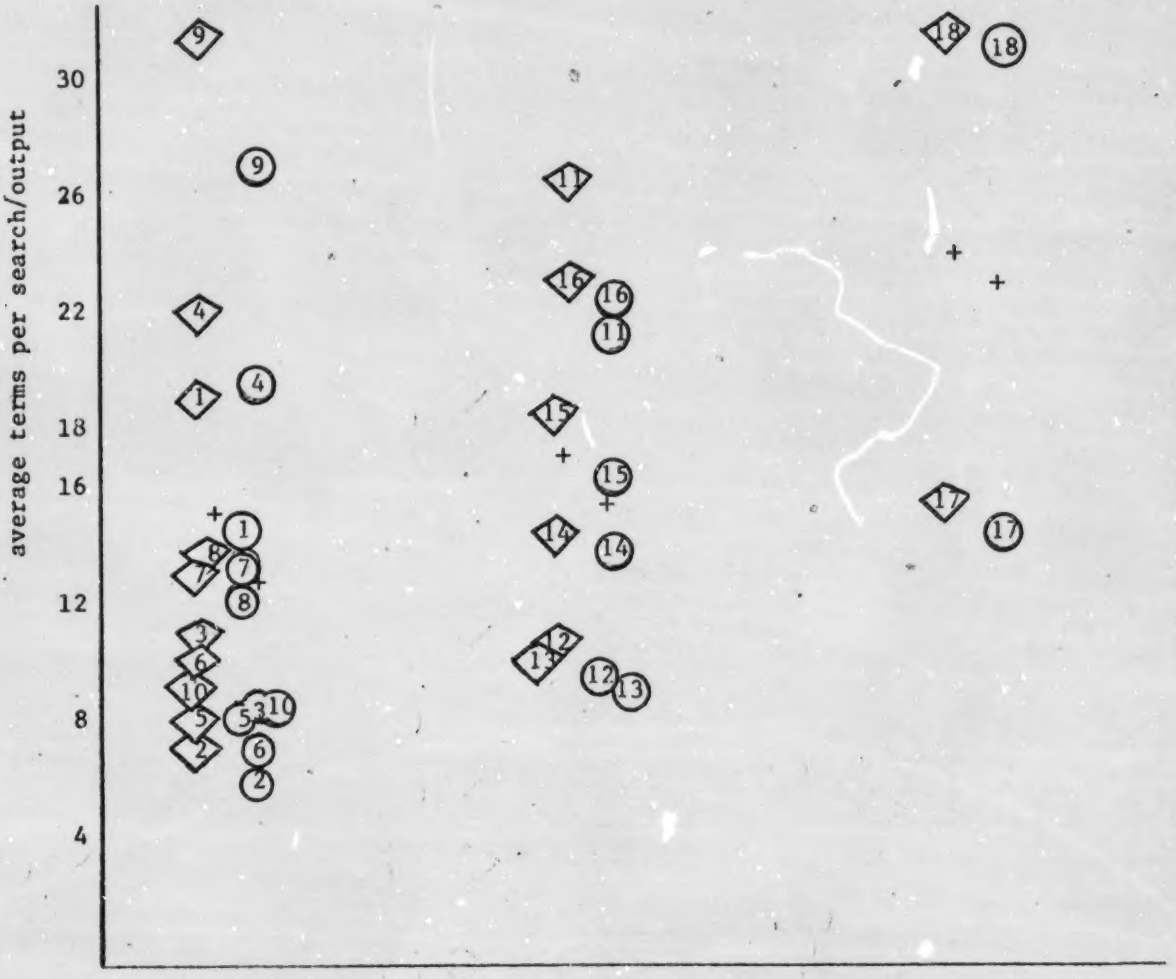


THREE TYPES OF ERIC ONLINE SEARCHERS

search (n) The numbers found on the graph refer to our numbering of the sets of traces received from 18 online searchers.

GRAPH 4 -- NUMBER OF SEARCH TERMS USED BY THREE TYPES OF SEARCHERS

In this illustration showing the average number of search terms among the three types of searchers, an increase can be detected from controlled vocabulary to free text searchers.



av = 14.6/12.5 (search/output)

controlled vocabulary searchers

av = 17.1/15.5 (search/output)

combination free text and controlled vocabulary searchers

av = 24.1/23.6 (search/output)

free text searchers

THREE TYPES OF ERIC ONLINE SEARCHERS

search ◇n

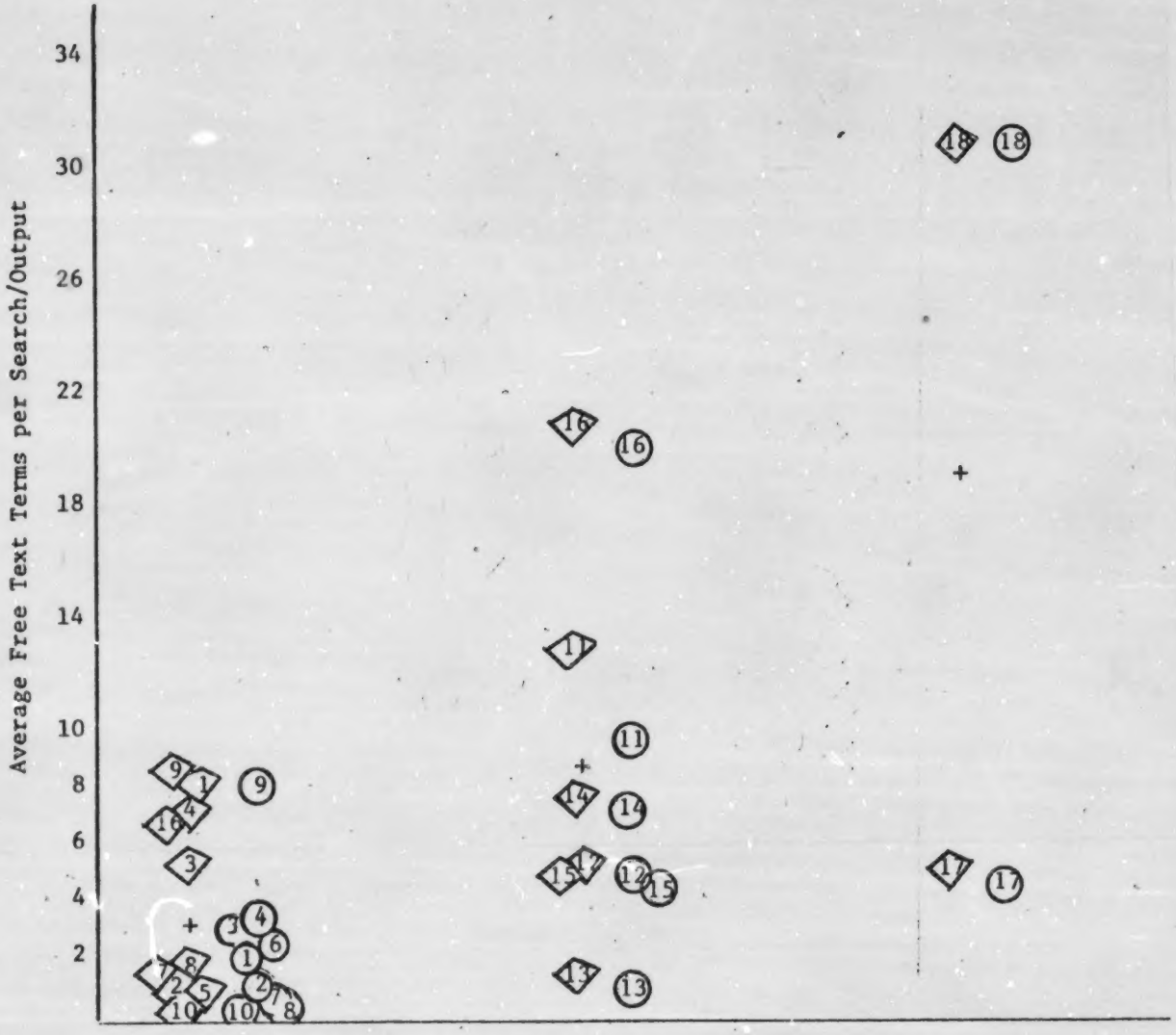
output (n)

average +

The numbers found on the graph refer to our numbering of the sets of traces received from 18 online searchers.

GRAPH 5 -- NUMBER OF FREE TEXT SEARCH/OUTPUT TERMS

One would expect then that the average number of free text terms per search increases from controlled vocabulary searchers to free text searchers.



THREE TYPES OF ERIC ONLINE SEARCHERS

(n) = search

(n) = output

The numbers found on the graph refer to our numbering of the sets of traces received from 18 online searchers.

SUMMARY AND RECOMMENDATIONS FOR FURTHER RESEARCH

In our discussions with online searchers, we found few differences among types of searchers in terms of sources for searching vocabulary, ERIC Thesaurus usage, or institutional affiliation. However, plotting averages and percentages of free text term usage according to searchers' responses to the question, "Do you perform primarily descriptor searches?", yields a distinct upward trend from controlled vocabulary to free text searchers. Unfortunately, only two of six free text searchers provided us with copies of their search traces, and it is obvious from the graphs that they vary considerably. A larger group of respondents in all categories, particularly free text, is suggested in future work.

In-depth studies might also point out why searchers chose to represent the output in a greater percentage of controlled vocabulary than the search. What we might speculate when reviewing the trace as factors influencing the searcher's decision to employ more controlled vocabulary in the output, ought to be matched with what individual searchers say about their own performance. In much the same way, we can report that searchers who limited by major descriptor maintained the limitation when ordering output. More outputs than searches are limited by major descriptor. In fact, one out of four searches contain more major than minor descriptors. Free text, combination, and controlled vocabulary searchers devised strategies composed solely of major descriptors. However, finding out from the searcher what factors could have influenced him to express the search in this way may isolate there with greater certainty. Let us speculate that final set size, precision, mapping of the controlled vocabulary onto concept groups of the query, or user objective, may in part contribute to major descriptor searches.

We have reported that one-third of ERIC searches are comprised of free

text, but this ratio varies among type of searchers. Further study might focus on some of the following aspects:

1. Searching style in other controlled vocabulary data bases, e.g., Psychological Abstracts, NTIS, ECER, Medline.

2. Using the search trace as a sort of stimulated recall, the searcher discusses what aspects of the online interaction influenced his decision to perform a task in a certain way.

3. Whether a significant distinction can be made between search and output, and what factors contribute to this distinction.

4. Is an ERIC free text searcher defined as one who uses 50% free text, 95%? Does he always input at least one statement of free text? Is free text searching an inclination to perform a certain way, or must it be accompanied by output results?

II. RETRIEVAL SYSTEM FEATURES

One-third of the 650+ search traces we examined employed some aspect of free text searching. We monitored all searches in order to count occurrences of retrieval system capabilities that accomodate free text searching (e.g., truncation, word proximity). Since the number of searches in which these features occurred was small, we were able to cover other capabilities (e.g., controlled vocabulary, truncation, online thesaurus and dictionary usage). It is noteworthy that no SDC search trace sets are examined, as one can see from Table 1.

One set of SDC traces was submitted to the Free Text Project. Unfortunately, it was unsuitable for our purposes because the online or offline print statements were cut off. Also, earlier portions of the search were cut off from the trace which the searcher referred to in the uncut strategy. This precluded counting controlled vocabulary and free text search terms per search and output.

A review of those search features which we monitored follows:

Table 4. Search Features of Retrieval Systems Reviewed by Project Staff

I. Free Text Searching

1. word adjacency, (W), ADJ
2. word proximity, (nW)
3. same field search, (F), SAME
4. sentence search, (S)
5. citation search, (C)
6. single terms (untruncated)
7. free text truncation, ?, $
8. fields specified for all of the above (I.1-I.7)
9. online dictionary

II. Controlled Vocabulary Searching

1. online thesaurus
2. controlled vocabulary truncation

III. "Housekeeping" Functions

1. execute save search
2. display sets

A. Word Adjacency

(Examples: HEAD ADJ START.TI,ID.; NEEDS(W)SENSING)

Searchers used this special feature 450 times. Specifically, all but one of the 18 respondents employed word proximity to search free text fields. Ninety percent of the word proximity statements were not limited by field, which resulted in retrieval from all basic index fields. When field delimiters occurred in word proximity statements, favored combinations were identifier, title, and descriptor fields, or title and identifier fields. Negative paragraph qualification was designated by one BRS search in order to exclude fields; in this case, the abstract.

B. Word Proximity/Sentence Searching

(Examples: GIFTED(3W)IDENTIFICATION; SPORTS WITH INJURIES)

This feature was reported less frequently than word adjacency. All six BRS searchers used sentence searching. On the other hand, less than half LRS searchers used word proximity (DIALOG's (nW) or (S)). One LRS searcher accounted for 105 of the 135 uses of the function among all respondents. In all but three of 135 occurrences, no field delimiters were employed.

C. Field Searching

(Examples: LASER SAME TECHNICIAN; COGNITIVE(F)MAPPING)

Only two of 13 LRS searchers contributed to the 23 occurrences in 156 LRS search traces. There were no instances of more than one (F) in a single statement, e.g., LIBRARY(F)COLLECTION(F)DEVELOPMENT. In only two of the 23 statements the searcher limited retrieval to title, identifier, and descriptor fields. None of the 472 BRS traces contained the SAME operator. It would appear that this operator will increase in usage as BRS unbinds descriptors so that searchers concerned about precision can qualify SAME statements by controlled vocabulary fields.

D. Single Free Text Terms

(Examples: WOMEN; MATHEMATICS/TI,DE,ID)

Next to word adjacency, single free text terms were reported most frequently (335 times). Over eight of every ten single terms was not limited by field. Favored combinations were identifier, title, and descriptor fields, or

in BRS, negative qualification of abstract, corporate source, and sponsoring agency fields.

E. Truncation

Single free text terms. Half of the respondents did not use truncation on a single free text term. Embedded truncation appeared once in 657 traces for the term WOM?N. Nineteen of the 90 truncated single terms employed single-character or multi-character trunction (e.g., OBSTACLE? ?, UNIVERSIT$3). It is important to note that truncated, single free text terms were limited by field (e.g., PRESIDENT$1..AB.) by only two of six BRS searchers. This capability is not available on LRS.

F. Controlled Vocabulary

It was not unusual to find BRS searchers truncating controlled vocabulary terms (e.g., EDUCATIONAL-A$) in order to retrieve multi-term descriptors like Educational Sociology, Educational Specifications, etc. This procedure is a short-cut to typing in individual descriptors. However, BRS retrieves only up to the first 100 entries in the dictionary file that begin with the input keyword or stem. One might find searchers entering series statements like EDUCATIONAL-A$, EDUCATIONAL-M$, EDUCATIONAL-R$, which resembles a sort of estimation by the searcher concerning when 100 entries have been retrieved. Although BRS plans to remedy this by unbinding descriptors and entering them into the basic index, some message ought to be included in the overflow statement (e.g., E147 TOO MANY WORDS FOR _____. QUERY ACCEPTED WITH WORDS AND COUNTS FOR TERM ONLY AS BEFORE THE OVERFLOW), which informs the searcher about the last term retrieved prior to the overflow. Then the searcher can continue by truncating from the overflow or aborting the strategy. Another alternative would be to report the overflow to the searcher and request whether he wishes to

continue from the point of overflow or abort -- similar to ORBIT's dialogue during a stringsearch.

G. Positional Operators and Truncation Searching

(Example: ROLE ADJ IDENTIFICATION$1; (ONE TWO THREE) ADJ YEAR$1 ADJ OLD$1

DIALOG does not yet offer text and truncation capabilities within a single search statement. All but one BRS searcher (who submitted one BRS search trace) used this feature. Some examples are as follows:

```
BASAL ADJ (READER$ READING)
(LEARNING COGNITIVE) ADJ STYLE$..AB.
TITLE ADJ (I ONE)
```

This feature is used primarily to capture variant word forms, especially plurals and synonyms with common stems. It is important to stress that BRS searchers used text and truncation searching 60 times in 472 searches. Four of six BRS searchers employed nesting (e.g., BASAL ADJ (READER$ READING)), but this accounts for no more than 28 search statements in 472 BRS searches.

H. Multiple Positional Operators

(Examples: WORK ADJ VALUES ADJ INVENTORY; SELF(W)FULFILLING(F)PROPHESY)

Designating more than one text searching delimiter (e.g., (F)(W)(nW)) was done infrequently by LRS searchers. Of three LRS searchers using this capability, one contributed 16 of the 18 usages in 156 LRS searches. All BRS searchers employed multiple text searching delimiters but proportionately less than LRS searchers. However, BRS allows positional and logical operators nested in a single statement (e.g., PARENT$2 ADJ (TRAINING OR EFFECTIVENESS) or AGE$ ADJ (THREE FOUR FIVE SIX)). Thus BRS searchers specify multiple positional or logical operators as infrequently as LRS searchers using multiple positional operators.

I. Online Dictionary

Only one BRS searcher submitted a display sets summary to us so that we

were unable to find out whether he used the ROOT command. However, if his use of BRS parallels that of the five other searchers, he did not call up the online dictionary. In 457 BRS search histories, the root command does not appear once. Perhaps, when provision is made for selecting terms from the ROOT display, searchers will be more inclined to expand terms. Five of 13 LRS searchers viewed and/or selected from the online dictionary. Only in one of 36 instances did the searcher limit the dictionary by field (YR = 76) rather than view the basic index display (if searcher entered a single term) or controlled vocabulary portion (if searcher entered a phrase).

J. Online Thesaurus

Six of 14 searchers having access to the ERIC Thesaurus online used the display during their search. Two of the six account for over 85% of its use.

K. Citation Searching
(Examples: TITLE(W)I(C)SPORTS)

This text searching delimiter, the most general of positional operators, was not found in any of the 156 LRS traces. As a single search statement, particularly when truncation is available, it allows the searcher to enter two or three heavily posted terms associated with the concept groups of a query and receive an estimate of final set size. This operator ought to become exceedingly useful when LRS allows truncation with positional operators. In this way the searcher may specify variant forms and plurals in one query.

L. Search Saves

Six of the 18 searchers used Search Saves. However, one searcher accounted for half the 83 reported saves. It is interesting to note that one search by this searcher was comprised of a single descriptor and six Search Saves. Search Saves used at this institution are prepared in advance on frequently-occurring facets such as Junior High, Special Education, Curriculum Guides, Pros and Cons.

An approach to formulating Search Saves without paying for storage costs was cited by one LRS searcher who called up heavily posted descriptors through the online thesaurus for frequently-occurring facets (e.g., Handicapped, Audio-visual Aids) and selected from the display according to a prescribed routine. Public Search Saves should alleviate much of the concern about storage costs, and when searchers become accustomed to entering them into the strategy, act as a promotion for private Saves and in-house or local user group sharing of Search Saves.

M. CONCLUSIONS ABOUT USE OF RETRIEVAL SYSTEM FEATURES

DIALOG's (C) and BRS' SAME features were not used. Fifty percent of Search Saves were accessed by one searcher. Two of six LRS searchers who called up the online thesaurus and selected term account for 85% of the reported usages. Field searching in LRS was done 23 times by only two of 13 searchers. Nesting occurs 24 times in 472 BRS searches. Negative paragraph qualification was employed by two BRS searchers, one of thom employed it 30 of the 32 reported usages. Display sets command was used by two of the 18 respondents. It is evident that these and other retrieval system features covered earlier are used infrequently, or by few of the total number of searchers. A partial explanation may be that these features accommodate free text searching, and since the majority of searches we reviewed were expressed in controlled vocabulary, these features are employed infrequently.

We might speculate then that the controlled vocabulary is adequate for representing most ERIC searches, so that only the most rudimentary features, selects, combines, and prints, are necessary. Retrieval system personnel have pondered whether searchers have developed sophisticated skills and techniques in order to accommodate enhanced searching capabilities. If searchers do not possess adequate skills, then a sophisticated system may only serve to encumber beginners, and confuse the majority. On the other hand, it may be reasoned that searchers employ

the fundamental techniques of the system which allow them to "get by" with representing the majority of searches expending a minimum of effort and imagination.

III. VOCABULARY STUDIES OF FREE TEXT SEARCHES

Our vocabulary studies focused on determining what proportion of search statements were incapable of being represented in ERIC's controlled vocabulary. In most of these statements, one or more retrieval system features whose usage was described above is used.

We took a random sample of 166 free text search statements from all 18 searchers. We sought information for each free text statement on the following questions:

1. Is it a descriptor (e.g., PERCEPTION or SELF(W)CONCEPT)? If not, then
2. Is it cross-referenced to an accepted term (e.g., ASSERTIVENESS(W)TRAINING -- use ASSERTIVENESS)? If not, then
3. Is it a variant of a descriptor (e.g., AUDIOVISUAL(W)AID, Descriptor = AUDIO VISUAL AIDS)? If not, then
4. Is it expressable using descriptors (e.g., SPORTS$1 WITH INJURIES, Descriptors = ATHLETICS(C)INJURIES/DE)? If not, then
5. Is it expressable as an identifier?
6. Is it only expressable as free text?

Our findings are as follows.

1. Is it expressable as a descriptor? Forty-five terms mapped onto single term or descriptor phrases. Postings ranged from 11 for BASEBALL to 8475 for MODELS. The newest descriptor which was input as free text was BASEBALL, entered in 1975. Below are listed 44 terms* as they were accompanied by descriptor postings (from ERIC Thesaurus, 7th edition, 1977) and free text postings (1966-Jan. 1979, LRS). For the six descriptors in Table 4 added after 1966, the date added as a descriptor to the Thesaurus is given in parentheses.

*The ranking by descriptor postings highlights the wide discrepancy with free text postings. See Table 2 to see where these free text terms were found and also see the Rotated Descriptor display in the Thesaurus to see where the single words appear in multi-word descriptors.

Table 5. Terms with Descriptor and Free Text Postings (ERIC on LRS)*

(Ranked by Number of Descriptor Postings)

	Free Text Statement	Descriptor Postings	Free Text Postings
1.	MODELS	8,475	17,169
2.	ACADEMIC(W) ACHIEVEMENT	7,919	10,325
3	EVALUATION$1	6,372	55,931
4.	SELF(W) CONCEPT	4,600	6,393
5.	ATTITUDE$1	4,115	39,443
6.	MATHEMATICS	2,828	13,759
7.	MOTIVATION	2,774	9,750
8.	DEVELOPING (W)NATIONS	2,667	3,340
9.	READING	2,539	27,509
10.	OBJECTIVES	2,310	36,559
11.	COMPOSITION	2,229	6,011
12.	PERCEPTION	2,109	10,529
13.	COLLECTIVE (W)BARGAINING	1,477	1,695
14.	PRINCIPALS	1,462	3,359
15.	SELF(W) ACTUALIZATION	1,441	1,801
16.	HIGH(W) SCHOOLS	1,411	7,243
17.	INTERACTION	1,402	9,888
18.	INSERVICE (W)EDUCATION	1,360	2,383
19.	MANPOWER(W) UTILIZATION	1,312	1,488
20.	STUDENT$1	1,242	115,061
21.	PARENT(W)ROLE	1,042	1,382
22.	BASIC(W)SKILLS	964	2,774
23.	TEACHER(W)AIDES	925	1,250
24.	BUDGETING	800	2,394
25.	TELECOMMUNI-CATION/TI,ID,DE (Jan 1970)	766	939
26.	CONFLICT	612	4,580
27.	HANDICAPPED	577	16,985
28.	STATISTICS/ DE,ID,TI	561	5,850
29.	LEARNING(W) LABORATORIES	502	665
30.	VISUAL(W)AIDS (Apr 1972)	459	1,247
31.	EMPATHY	351	698
32.	LATIN(W)AMERI-CAN(W)CULTURE	350	400
33.	TRUSTEES.MJ, MN,TI,ID	317	1,061
34.	GERIATRICS (Aug 1968)	195	244
35.	TAPE(W)RECORDERS	195	234
36.	SEMANTIC(W) DIFFERENTIAL (Oct 1972)	169	704
37.	FRESHMEN	128	2,654
38.	SELF(W)CONGRU-ENCE	114	129
39.	COEDUCATION (May 1969)	111	171
40.	HOUSEWIVES	100	226
41.	ORIENTATION (W)MATERIALS	93	135
42.	GYMNASIUMS	75	109
43.	REMEDIAL(W) ARITHMETIC	29	42
44.	BASEBALL (Jun 1975)	11	76

*See Table 2 to see where single words may be found in free text searches.

2. Is it cross-referenced to an accepted term? Ten free text statements were cross-referenced to a descriptor. Searchers who used this tactic may have been including synonymous terms in order to enhance recall. The most recent cross-referenced term, Assertiveness, was integrated into the controlled vocabulary in 1977. In all ten instances, the searcher entered the accepted controlled vocabulary term, either as free text or controlled vocabulary. The ten terms and cross references are listed below.

Table 6. Lead-in Term and Descriptor Postings for Ten Terms

Lead-in Term Free Text Statement	Free Text Postings	Descriptor	Entry Date	Descriptor Postings (as of 4/79)
1. ASSERTIVE(W)TRAINING	57	Assertiveness	1977	59
2. AGED	1810	Older Adults	1966	1577
3. BEHAVIOR(ADJ)MODIFICATION	888	Behavior Change	1966	4502
4. GIRL$1	2952	Females	1966	5402
5. INCENTIVES	571	Motivation	1966	3185
6. MECHANIZATION/DE,TI,ID	164	Automation	1966	863
7. PARENT ADJ STUDY ADJ GROUP$1	0	Parent Conferences	1966	66
8. PREJUDICE	564	Bias	1969	942
9. SPORTS/DE,TI,ID	1101	Athletics	1966	1297
10. WOMEN/DE,TI,ID	6792	Females	1966	5402

3. Is it a variant of a descriptor? The 21 free text statements included in this category were either singular forms of plural descriptors, or portions of multi-term descriptors. At this point, it became necessary to scan the trace for terms synonymous to the free text statement in order to determine whether the searcher was interested in retrieving the entire descriptor. For example, a searcher inputting REPETITIVE may wish to retrieve the controlled vocabulary term REPETITIVE FILM SHOWINGS, but if synonymous terms (e.g., REPEATED, REPETITION) occur in the same trace, the term was not counted as a variant descriptor.

Table 7. Free Text Statements with Variant Descriptors.

	Free Text Statement	Variant Descriptors Possible to Retrieve
1.	PERCEPTION	part of 5 descriptors
2.	PREVOCATIONAL	Prevocational Education
3.	CONFLICT	part of 4 descriptors
4.	COMPETENCY	National Competency Tests
5.	BOOK/DE,TI,ID	part of 4 descriptors
6.	RURAL/DE,TI,ID	part of 18 descriptors
7.	COMPOSITION	part of 4 descriptors
8.	SIZE	part of 4 descriptors
9.	REQUIREMENTS	part of 4 descriptors
10.	HANDICAPPED	part of 23 descriptors
11.	PREREADING	Prereading Experience
12.	ALTERNATIVE	Alternative Schools
13.	MOTIVATION	part of 5 descriptors
14.	READING	part of 60 descriptors
15.	PRESCHOOL$	part of 10 descriptors
16.	AUDIOVISUAL(W)AID	Audiovisual Aids
17.	VIDEO(W)CASSETTE	Video Cassette Systems
18.	FAMILY(W)RELATIONSHIPS	Family Relationship
19.	JUNIOR(W)HIGH	part of 2 descriptors
20.	PROGRAM	part of 19 descriptors
21.	CATHOLIC	part of 5 descriptors

4. Is the statement expressable using descriptors? We have extricated statements from their original context of the search. Treating each statement as a single search query or facet of a query, we have found 24 instances in which the controlled vocabulary seems suitable. The table below lists the 24 free text statements entered by the searcher and our recommendations for possible descriptors. These are accompanied by controlled vocabulary terms input by the searcher in the same search in which the free text statement occurred. In two-thirds of the instances below, the searcher has selected both free text and controlled vocabulary terms to express the concept. In ten search statements in Table 8, the searcher missed items because their choice of free text and descriptors did not match possible descriptors (see 1, 3, 5, 6, 7, 8, 14, 15, 20, 24).

Table 8 : Free Text Statement in Search with Possible Descriptors.
(In search and not in search)

	Free Text Statement	Possible Descriptors	Descriptors Input by Searcher in Same Search as Free Text Statement
1.	COLLECTION ADJ DEVELOPMENT	Library Acquisition, Library Material Selection	Library Acquisition
2.	CITIZENSHIP ADJ EDUCATION	Citizenship	Citizenship.MJ,MN.
3.	PHYSICAL NEEDS	Health Needs or Physical Health	---
4.	LEARNING STATION	Learning Laboratories	Learning Laboratories, Centers of Interest, Skill Centers
5.	VISUAL PROCESSING	Visual Perception or Visual Learning	---
6.	NON(W)SEXIST	Sex Discrimination	---
7.	THREE ADJ YEAR$ ADJ OLD $2	Preschool Children	Early Childhood$, Early Experience, Nursery Schools
8.	SPORT$1 WITH INJURY$	Athletics/DF and Injuries/DF	Athletics.MJ,MN., Legal Problems, Laws.MJ,MN., Legal Responsibility
9.	SEVEN ADJ YEAR$ ADJ OLD $1	Grade 2	Grade 2, Elementary Education, Elementary Grades
10.	AGE ADJ THREE	Preschool Children	Preschool-$
11.	COMPENTENCY(W) REQUIREMENTS	Performance Based Education	Performance Based Education/ DE*, Graduation Requirements/ DE*, ID*
12.	LICENSES	Certification	Certification.MJ,MN., Teacher Certification
13.	MARINE ADJ SCIENCE	Marine Biology	Marine Biology
14.	YOUNG ADJ MEN/TI	Young Adults and Males/DF	Young Adj Men/ID
15.	SEXIST ADJ LANGUAGE	Language/DF and Sex Discrimination	Feminism.MJ,MN., Women's Studies, Language.MJ,MN.
16.	READING(W)LABORATORIES	Reading/DF and Learning Laboratories	Remedial Reading, Remedial Reading Clinics
17.	AGRICULTURAL CHEMISTRY	Agriculture/DF and Chemistry/DF	expanded 5 Agriculture, Agricultural Chemistry, and Agronomy descriptors and selected 41 descriptors from online Thesaurus displays

Table 8 (cont.)

	Free Text Statements	Possible Descriptors	Descriptors Input by Searcher in Same Search as Free Text Statement
18.	THEORY(F)ATHLETICS	Theory/DF and Athletics/DF	---
19.	MENTAL(W)HEALTH(1W) TEACHERS	Mental Health and Teachers/DF	Teachers/DE,ID, Mental Health
20.	BLACK(F)SPORTS	Blacks/DF and Athletics/DF	---
21.	SOCIOLOGY(F)ATHLETICS	Sociology/DF and Athletics/DF	---
22.	MINORITIES(F)ATHLETICS	Minority Groups and Athletics/DF	---
23.	TEACHERS(3W)ANXIETY	Anxiety/DF and Teachers/DF	Anxiety/DE,ID, Teachers/DE,ID
24.	COLOR(F)ATHLETICS	Blacks/DF and Athletes/DF	---

5. Is it expressable as an identifier, or if not,

6. Is it expressable as a free text statement? Descriptors can be used to represent free text statements in 99 of 166 instances. We then input the statement into LRS as an identifier and as free text, and noted our results.

Table 9: Free Text Statements with Postings in Identifier Field and other Free Text Fields.

	Free Text Statement	Postings: Identifier	Postings: Free Text
1.	MOLDING(W)FABRICATION	0	0
2.	CYCLICAL ADJ PROGRAM$	0	0
3.	FISCAL ADJ AUTONOMY	0	1
4.	STOPOUT	0	1
5.	VITA ADJ LITE	0	1
6.	NONTRADITIONAL(W)JOB	0	2
7.	AGRICULTURAL(W)CHEMISTRY	0	2
8.	PICTURE(W)LABELING	0	2
9.	NEEDS(W)SENSING	0	3
10.	AGGRESSION ADJ SCALE	0	4
11.	PREENGINEERING	0	4

(cont.)	Free Text Statement	Postings: Identifier	Postings: Free Text
12.	RESUME ADJ PREPARATION	0	7
13.	COMMISSION WITH QUALITY WITH FINANCING WITH COST	0	9
14.	LEFT(W)BRAIN	0	9
15.	PSYCHOLOGICAL(W)MODELS	0	9
16.	NATIONAL(W)CRIMINAL	0	13
17.	SOCIAL ADJ EDUCATORS	0	23
18.	ROLE(W)DIFFERENTIATION	0	30
19.	SHY	0	38
20.	NON(W)SEXIST	0	51
21.	WRITTEN(W)WORD	0	54
22.	WITHDRAWN	0	195
23.	LASER(F)TECHNICIANS	1	2
24.	SURVIVAL(W)TRAINING	1	8
25.	WORK(3W)INCENTIVES	1	24
26.	YOUNG ADJ MEN	1	42
27.	INQUIRY ADJ SCHOOL	2	13
28.	RETICENT	2	16
29.	SYMPATHY	2	58
30.	DISPLACED	2	99
31.	LIMITATION	2	216
32.	TM	2	1227
33.	ORGANIZATIONAL ADJ BEHAVIOR	4	64
34.	PRECALCULUS	5	9
35.	ENGLISH ADJ PRIMARY	5	42
36.	SIGNIFICANT(W)OTHERS	5	97
37.	ELIM?	6	2067
38.	SHYNESS	8	20
39.	IEP	8	57
40.	RETICENCE	9	23
41.	ANOMIA	9	33
42.	PREREADING	9	263
43.	RESERVE ADJ OFFICER ADJ TRAINING ADJ CORPS	10	25
44.	AUDIOTUTORIAL	11	62
45.	TELEGRAPH	12	64

(cont.)	Free Text Statements	Postings: Identifier	Postings: Free Text
46.	ADAPTIVE ADJ BEHAVIOR ADJ SCALE	13	24
47.	STOCHASTIC	13	60
48.	SUSTAINED ADJ SILENT ADJ READING	14	26
49.	ST ADJ LAWRENCE	14	47
50.	CLASS(W)INCLUSION	14	62
51.	WORK(W)VALUES(W)INVENTORY	17	28
52.	INFANT ADJ SCHOOL	18	74
53.	RETRIEVAL(F)PROBLEM	18	89
54.	LINKING	23	410
55.	PSSC	25	69
56.	EXEMPLARY	40	1029
57.	DISTAR	43	68
58.	SELF(W)PACED(W)INSTRUCTION	45	106
59.	FIELD(W)DEPENDENCE	47	106
60.	CORNELL	68	612
61.	INDONESIA	90	190
62.	HEADSTART	91	201
63.	DELPHI(W)TECHNIQUE	126	181
64.	EFFECTIVE	138	15190
65.	TITLE(W)I	1194	1521
66.	GREAT ADJ BRITAIN	1866	2124
67.	CALIFORNIA	3893	12024

Summary of Findings About Vocabulary Usage

Twenty-two statements, 14% of the sample, retrieved no citations as identifiers and were incapable of being expressed as descriptors. In conversation with online searchers, five types of questions may necessitate free text searching:

1. geographic areas
2. recent topics in education
3. specific named objectives
4. value judgments
5. action statements

Reviewing the list of free text statements, we see that:

1. Geographic areas account for four easily identifiable items.

2. Recent topics include such terms as class inclusion, fiscal autonomy, displaced homemakers, non sexist, traditional jobs, and stopout. This category is made up of catch words and familiar phrases.

3. Specific named objects may be comprised of:

   a. tests and rating scales: aggression scale, adaptive behavior scale, work values inventory;
   b. educational programs: headstart, sustained silent reading, precalculus, agricultural, chemistry, and ROTC.

   Problems sometimes arise when program names (e.g., USSR, TM, ESP) match acronyms used across many fields. Some provision ought to be made for topics in education and named objects in foreign countries: inquiry school or infant school.

4. Value judgments might include terms like: effective, validated, exemplary.

5. Action statements show relationships among terms or limit the scope of an aspect (e.g., eliminate, limitation, linking, and repetitive).

We might add another category for topics marginal to the scope of education (e.g., left brain, telegraph). Individual or psychological characteristics (e.g., shyness, sympathy, reticence) might make up a separate category. The original list has now been amended to contain:

1. geographic areas

2. recent topics
   a. in education
   b. marginal to the scope of education
   c. specific to foreign countries

3. specific named objects
   a. in education
      1. tests and rating scales
      2. programs
   b. marginal to education
   c. specific to foreign countries

4. value judgments

5. action statements

6. individual or psychological characteristics

Undoubtedly, additional categories could be added to the list. We have attempted to provide some indication concerning what types of queries are input as free text. Further work in this area may entail studying free text statements in the context of the search. In our study, free text and controlled vocabulary searchers differed little in the types of free text statements they entered. All three groups of searchers were inclined to represent descriptors or descriptor-based queries in free text. Unexpectedly, statements like EDUCATIONAL(W)OBJECTIVES and SELF(W)CONCEPT appeared in traces of controlled vocabulary searchers. Two-thirds of the free text terms which were unable to be expressed by descriptors had fewer than 15 identifier postings. Only two free text statements yielded zero postings for both identifier and free text searches. For the entire sample, one out of every eight statements could not be represented in the controlled vocabulary. This ratio shrinks considerably when the free text statement is treated in the search where two out of three terms are descriptors. The following diagram shows what portion of an ERIC search is unable to be expressed in the controlled vocabulary based on our random sample. If it is true then that over 95% of an ERIC search can be described by the controlled vocabulary, we might take a closer look at aspects like type of searcher, institutional affiliation, and user objectives, in order to discover relationships between these and free text searching. See Figure 1 on next page.

These findings point to the need for a new online aid for ERIC searchers, namely, a linkage between their "free text" input and the controlled vocabulary -- something like the lead-in terms, the use references in the printed thesaurus. Such an endeavor would greatly enhance the online searcher's output. A post hoc analysis of searches, such as that presented here, could upgrade the new online thesaurus very quickly.

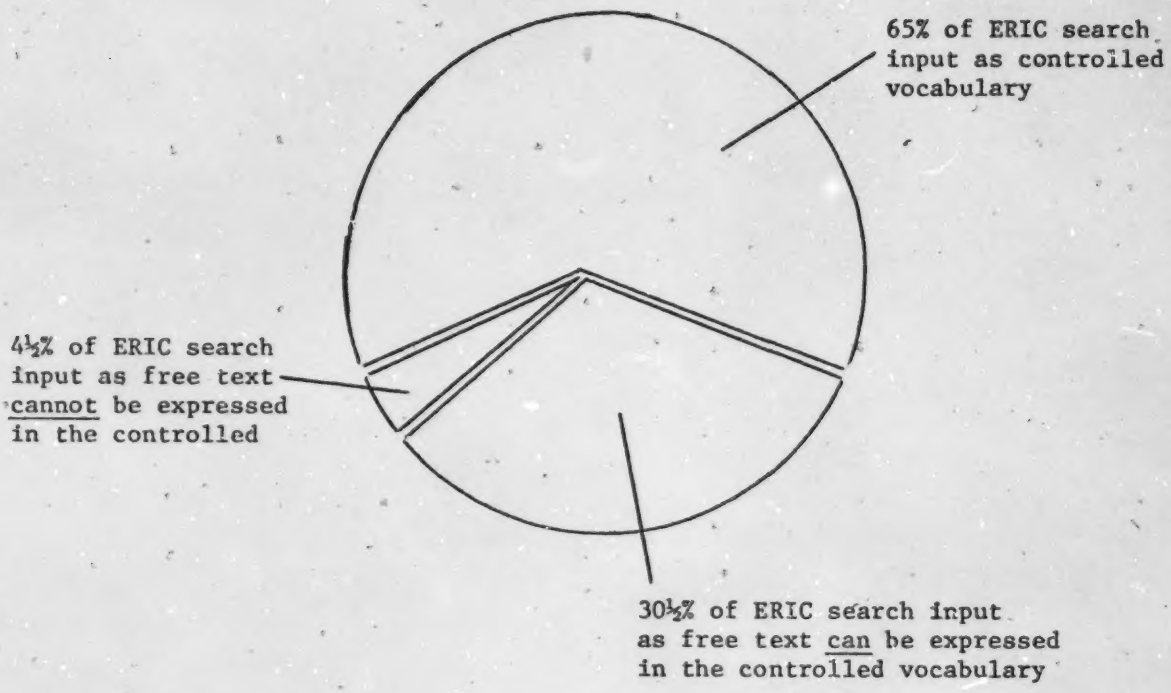

65% of ERIC search input as controlled vocabulary
4½% of ERIC search input as free text cannot be expressed in the controlled
30½% of ERIC search input as free text can be expressed in the controlled vocabulary

Figure 1. ERIC Online Searches -- Percentage of Input as Free or Controlled Vocabulary

ERIC/IR
Special Project Report

ONLINE SEARCHING OF ERIC

PART III

ONLINE SEARCHING TESTS

ERIC Clearinghouse on Information Resources
Syracuse University
1979

PART III

ONLINE SEARCHING TESTS

by Karen Markey and Pauline Atherton

Abstract

Six search queries from the ONTAP: Online Training and Practice Manual for ERIC Data Base Searchers (ED 160 108) are used to test controlled and free text vocabulary searching strategies. Recall and precision results were compared. Failure analyses were included to diagnose why a search strategy resulted in low recall, low precision, or missed relevant items. Overall, the free text search formulations had higher recall and the controlled vocabulary formulations had higher precision. We concluded that free text searching can often be the better search formulation.

Note:

We would like to express our appreciation for the cooperation we received during the course of this particular project from staff members at several ERIC Clearinghouses:

Sydney Meredith/SO
Pat Spellman/SP
Mike Butler/SP
Norma Howard/PS
Dorothy Bloch/EC
Lynn Smart/EC
Lynn Haupt/HE
Barbara Booth/JC

(See Contents on next page)

PART III

ONLINE SEARCHING TESTS

Contents

	Page
Introduction	1
Search Topic #1 - Identification of the Gifted	5
Search Topic #2 - Higher Education Enrollment Projections For the Near Future (1975-95)	9
Search Topic #3 - Junior College Enrollment Projections For the Near Future (1975-95)	13
Search Topic #4 - Drug Abuse, Including Alcohol, Among Students of Elementary and Secondary Schools	16
Search Topic #5 - Use of School Busing to Achieve Racial Integration	21
Search Topic #6 - Jean Piaget's Theories, and Thought Processes or Language Development of Children	23
Observations and Conclusions	29
(A) Relevance Assessments	29
(B) Recall/Precision Performance	29
(C) Overlap	31
Recommendations	32
Appendix to Part III	33

Tables

		Page
Table 1.	ERIC Unit Record-BRS/LRS/SDC Search Qualifiers	3
Table 2.	Precision and Recall of Online Searching (Summary)	29
Table 3.	Recall and Precision for Six Searches	30
Table 4.	Overlap of Items Retrieved by Controlled and Free Text Searches	31

ONLINE SEARCHING TESTS

INTRODUCTION

Search formulations online can be expressed in controlled vocabulary and in free text. We explored several search topics used in the ERIC ONTAP data base* to test which search strategy was more successful. Since the ERIC ONTAP file was specifically designed to allow online searchers to practice and compare their search results with the results of a "perfect answer set," we thought this would serve as a useful test environment for purposes of comparing several search formulations.

The ERIC ONTAP Manual* has the free text searches. Staff at several ERIC Clearinghouses cooperated with us to compare free text search results for six search topics in ONTAP with a search formulation composed exclusively in controlled vocabulary. They also helped us assess the output for relevance.

Our report contains:

1. a discussion of the controlled vocabulary and free text formulations;
2. the search results measured by recall and precision;

---

*The ERIC ONTAP data base was developed by Charles Bourne and his colleagues at the Institute for Library Research (ILR) at the University of California, Berkeley, with the support and encouragement of Oscar Firschein of Lockheed Information Systems. This file (File 201 on LRS/DIALOG), made up of 1975 ERIC (CIJE and RIE) records, is specially designed to permit searchers to compare their search results with the results of a "perfect search." The Berkeley group developed "answer sets" for some 20 search topics by exhaustively searching the ONTAP data base for relevant hits. The composite search output of all their strategies was assessed for relevancy. Thus the "answer set" contains only items judged as being relevant to the topic. The ILR group offered a representative strategy specifically constructed to retrieve all the items in the test set, i.e., one developed for high recall. This formulation, a composite which was assembled with hindsight after viewing the results of several formulations, features free text. From a draft report prepared by Charles Bourne and his colleagues at ERIC/IR we worked to produce an online training and practice manual for ERIC searchers. Karen Markey and Pauline Atherton worked on this manual. (ED 160 108) The manual contains several searches which are not, at present, in File 201 on LRS/DIALOG. In this report, we have used three searches in File 201 and the manual, and three only in the manual.

3. the overlap in search results; and

4. comments about conflicting relevance assessments.

We included failure analysis to diagnose why a search strategy resulted in low recall, precision, or missed relevant items. The six search topics we chose were:

1. Identification of the Gifted

2. Higher Educational Enrollment Projections for the Near Future (1975-95)

3. Junior College Enrollment Projections for the Near Future (1975-95)

4. Drug Abuse, Including Alcohol, Among Students of Elementary and Secondary Schools

5. Use of School Busing to Achieve Racial Integration

6. Jean Piaget's Theories, and Thought Processes or Language Development of Children

Three of these search topics (#4-6) are questions in DIALOG's ONTAP® ERIC (file 201). The remaining three (#1-3) were generated specially for the ERIC ONTAP Manual. (see ED 160 108)

The ERIC Clearinghouses working with us were:

1. Handicapped and Gifted Children
2. Higher Education
3. Junior Colleges
4. Teacher Education
5. Social Science/Social Studies Education
6. Early Childhood Education

To measure recall we had to find in the ERIC ONTAP test set all the relevant items on a given topic. To do this, it was necessary to search all subject-rich fields, title, identifier, descriptor, abstract, corporate source, sponsoring agency, and note fields. When relevant judgments were made, they were not restricted to an evaluation of the controlled vocabulary fields of

Table 1. ERIC UNIT RECORD - BRS/LRS/SDC Search Qualifiers

ERIC Field Names*	ERIC Codes (RIE/CIJE)	BRS-Search Qualifier	LRS-Search Qualifier	SDC-Search Qualifier
Accession Number	0010	.AN.	---	/AN
Clearinghouse Accession Number	CH 0011	---	---	/CHAN
Clearinghouse Code		.CH.	CH=	/CC
Title*	001A	.TI.	/TI	/TI
Author*	AUTH 001B	.AU.	AU=	/AU
Corporate Source*	INST 001C	.IN.	/CS	/OS
Sponsoring Agency	SPON 0020	.SN.	/SA	/SPO
Descriptor*	DESC 0023	.MJ,MN.	/DE or /DF or /DE* or /DF*	/IT
Identifier*	IDEN 0024	.ID.	/ID or /IF or /ID* or /IF*	/IT
Abstract*	ABST 002C	.AB.	/AB	/BI
Descriptive Note	NOTE 0026	---	/NT	---
Publication Year*	PDAT 0017	.YR.	YR=	FROM__ TO__
Pagination*	PAGE			
Journal Name*	JNL 0032	---	JO=	/JC
Legislative Authority Code		.PA.	AC=	/LPA
Contract/Grant Number	GR 002F CONT 002E	---	CN=	/NU
Bureau/Project Number	BN 0030	---	PN=	/NU
Report Number*	REPNO 002D	---	RN=	/NU
Corporate Source Code	0080	.IN.	SC=	---
Sponsoring Agency Code	0084	.SN.	SC=	---
Document Type*	PUBTYPE 0013	.PT.	DT=	/DT
Group Code	0014(?)	---	GC=	---
EDRS Price	0025	---	---	---
Availability Level	AVAIL 0031	---	---	---
Update	0001	---	UD=	/UP
Issue	002B	---	IS=	/IS

*Several of the fields in the ERIC record are listed in the "Recommended Interchange Specifications for Computer-Readable Bibliographic Data Bases" (April 1978 publication of ASIDIC-EUSIDIC-ICSU-AB-NFAIS). Those found in their list of essential data elements are marked here with an asterisk (*). Their field names vary from ERIC, e.g., Journal is Serial Title, Title is Title of contribution, Publication Date is Date of issue or imprint, etc.

As of January, 1979, ERIC added three new fields (GOV, LANG, and GEO) to the RIE & CIJE Master Tape. These are not in this table because the respective systems have not yet indicated whether or not they will be searchable directly or by prefix codes.

each retrieval item. We considered all fields of the ERIC unit record. (See Table 1.)

The high recall formulations recognize and employ the range of search features of the DIALOG retrieval system and its manipulation of all subject-bearing fields of the ERIC unit record.

Each of the six ERIC Clearinghouses was sent four items.

1. the relevant output of the high recall search (this corresponds to the ONTAP "answer set" for their Clearinghouse search topic);
2. the printout of the online search we performed using controlled vocabulary formulations;
3. the total output of the controlled vocabulary formulation; and
4. the listing of ERIC accession numbers of both formulations, organized in numerical sequence and showing overlap between search formulations.

We asked each Clearinghouse to review the four items to determine relevant citations for the topic, and derive an optimum formulation using controlled vocabulary. If they felt that our controlled vocabulary formulation was not suitable for expressing the search, they devised another strategy and assessed the relevance of its output.

* * *

We first present each of the six search topics with the two formulations, the search results, and an analytic statement about the results.

After this search-by-search analysis, we have a summary and some conclusions.

The free text search is a composite search, developed after several tries (see footnote on page 1 of this report) to achieve high recall. The controlled vocabulary search strategy was developed by the ERIC Clearinghouse staff members who cooperated with us. We did not revise or amplify the strategy in any way.

SEARCH TOPIC #1 (#11 in ERIC ONTAP Manual)*
(not in LRS File 201)

Identification of the Gifted

SEARCH FORMULATION --- The free text search is a composite search developed for high recall (see footnote on page 1 of this report.) An ERIC Clearinghouse staff person contributed the controlled vocabulary search

In Free Text			In Controlled Vocabulary		
1	114	GIFTED	1	87	GIFTED/DE
2	113	TALENT?	2	136	IDENTIFICATION/DF
3	2262	IDENTIF?	3	4	IDENTIFICATION TESTS
4	45	(1 OR 2) AND 3	4	7	TALENT IDENTIFICATION
			5	15	ABILITY IDENTIFICATION
			6	157	2-5/OR
			7	17	6 AND 1

STATISTICS ---

	Free Text	Controlled Vocabulary
Total Relevant = 27		
Total Retrieved	45	17
Total Relevant Retrieved	27	15
Total Relevant Missed	0	11
RECALL	$\frac{27}{27} = 100\%$	$\frac{15}{27} = 56\%$
PRECISION	$\frac{27}{45} = 60\%$	$\frac{15}{17} = 88\%$

ANALYSIS ---

The free text search appears to have better results. Of the six controlled vocabulary formulations this is the lowest recall score recorded. The searcher has chosen to select a single descriptor to represent the facet Gifted. This term is combined with four full descriptors for the concept Identification. The following list cites documents not retrieved by the con-

*See ED 160 108

trolled vocabulary formulation and the locations of portions of text relevant to the topic. In two instances no descriptors are assigned which refer directly to Identification or Gifted. All other indexing records contain descriptors for either facet, Gifted or Identification.

(NOTE: One record, ED 110 160, retrieved by both strategies and previously assessed as non-relevant by ERIC/IR, should be included in the ERIC ONTAP answer set as a relevant item.)

Reviewing the following list demonstrates that descriptors alone are not suitable for achieving high recall in this topic.

ED 100 863 Early Childhood Identification (title), ". . . on the early childhood identification of the handicapped or gifted . . ." (abstract)

ED 107 746 ". . . with a method for locating high ability inner-city students . . ." (abstract), *Educational Diagnosis, Academic Ability

ED 106 140 Experimentation for the Mathematically Talented (title), "identification of qualified students through difficult tests" (abstract), *Gifted

ED 105 329 ". . . to identify and counsel . . . students who are mathematically gifted" (abstract), *Gifted

ED 104 971 ". . . seeks to identify and encourage academically talented students" (abstract), *Talented Students

ED 102 771
ED 102 772 "Policies and regulations governing programs for . . . gifted and talented . . ." (abstract), Identification (minor descriptor)

ED 100 102 ". . . and identification and selection of gifted children" (abstract), *Gifted

ED 100 072 "Educational programs for Gifted Pupils" (title), "chapters dealing with the following topics: identification of pupils" (abstract), *Gifted

ED 097 788 "Programs serving the following: . . . gifted" (abstract), Identification (minor descriptor)

Some use of title searching may be valuable in order to retrieve additional

relevant items while not sacrificing precision; but, as shown by the free text performance, low precision may result when searching extends over all subject-bearing fields.

The free text search resulted in 100% recall and 60% precision. We excerpt here from the ONTAP Manual to show how this search formulation developed online from the use of the citation pearl growing approach on to a "tight" free text search.

*Step 6: Conceptualizing the Search as Input to the Retrieval System

Using the citation pearl growing approach, the searcher begins with a very direct search on the most specific search terms. Expressing the topic in the terms used by the requester, identification and gifted, is limited to retrieving citations whose titles contain these terms. The searcher then prints a few citations in format #2, checking the descriptor field in order to gather other search terms.

Formulation #1	Comments
1 S GIFTED(F)IDENTIFICATION/TI	Five citations are retrieved; the searcher, confident that all five are relevant, checks the index terms in order to add more search terms to the formulation.

TYPE 1/2/1-2

EJ103421 UD503083
THE IDENTIFICATION OF MENTALLY GIFTED, "DISADVANTAGED" STUDENTS AT THE EIGHTH GRADE LEVEL
FITZ-GIBBON, CAROL T.
JOURNAL OF NEGRO EDUCATION, 43, 1, 53-66 W74
*DESCRIPTORS: *JUNIOR HIGH SCHOOL STUDENTS/ *TALENT IDENTIFICATION/ *DISADVANTAGED YOUTH/ *URBAN SCHOOLS/ NEGRO STUDENTS/ GIFTED/ TALENTED STUDENTS/ ECONOMICALLY DISADVANTAGED/ PREDICTIVE MEASUREMENT*
*IDENTIFIERS: *CALIFORNIA*

---

*From Markey, Karen and Pauline Atherton. ONTAP: Online Training and Practice Manual for ERIC Data Base Searchers. (Syracuse, NY: ERIC Clearinghouse on Information Resources, Syracuse University, June 1978), pp. 118-120. ED 160 108.

ED108278 EA007191
GIFTED STUDENTS: IDENTIFICATION TECHNIQUES AND PROGRAM ORGANIZATION
DOOB, HEATHER S.
EDUCATIONAL RESEARCH SERVICE, WASHINGTON, D.C.
APR 75 61P.
AVAILABLE FROM: EDUCATIONAL RESEARCH SERVICE, INC., 1815 NORTH FORT MYER DRIVE, ARLINGTON, VIRGINIA 22209 ($5.00, PAYMENT MUST ACCOMPANY ORDERS $10.00 OR LESS)
DOCUMENT NOT AVAILABLE FROM EDRS
*DESCRIPTORS: ABILITY IDENTIFICATION/ BIBLIOGRAPHIES/ ELEMENTARY SECONDARY EDUCATION/ *GIFTED/ *PROGRAM CONTENT/ *PROGRAM DESCRIPTIONS/ *SUPERIOR STUDENTS/ *TALENTED STUDENTS/ TALENT IDENTIFICATION*

Gathering the term "Talent" from the descriptor field, the searcher revises the strategy in the second formulation restricting all search terms to occurrences in the descriptor, identifier or title fields.

Formulation #2

1 S GIFTED/DE,ID,TI
2 S TALENT/DE,ID,TI
3 S TALENTED/DE,ID,TI
4 S IDENTIFICATION/DE,ID,TI
5 C (1 OR 2 OR 3) AND 4

Employing the same search terms as formulation #2, the following high recall formulation searches all subject-conveying fields rather than being limited to the title and assigned index term fields.

Formulation #3	Comments
1 S GIFTED?	Included DESCRIPTORS: GIFTED, IDENTIFICATION ABILITY IDENTIFICATION, 4 multi-term IDENTIFICATION DESCRIPTORS, TALENT IDENTIFICATION TALENT, 3 multi-term TALENT DESCRIPTORS, TALENTED STUDENTS
2 S TALENT?	
3 S IDENTIF?	
4 C (1 OR 2) AND 3	

The Briefsearch, a single search statement, resembles the wording of the search topic.

Briefsearch

1 S IDENTIFICATION(F)GIFTED

(Reader: Keeping in mind that high precision is a search objective as well as delivering output to the requester as soon as possible, which formulation above fulfills both objectives?_____)

SEARCH TOPIC #2 (#12 in ERIC ONTAP Manual)*
(not in LRS File 201)
Higher Education Enrollment Projections
For the Near Future (1975-95)

(only journal citations, i.e., EJs)

SEARCH FORMULATION --- The free text search is a composite search developed for high recall (see footnote on page 1 of this report). An ERIC Clearinghouse staff person contributed the controlled vocabulary search.

In Free Text			In Controlled Vocabulary		
1	832	ENROLLMENT?	1	4875	HIGHER EDUCATION
2	948	PREDICT?	2	669	POSTSECONDARY EDUCATION
3	302	PROJECTION?	3	3130	COLLEGE/DE
4	90	FORECAST?	4	1527	COLLEGES/DE
5	1750	FUTUR?	5	584	UNIVERSITIES/DE
6	172	1 AND (2 OR 3 OR 4 OR 5)	6	273	UNIVERSITY/DE
7	9774	.EXECUTE Search Save for HIGHER EDUCATION (see Appendix)	7	176	ENROLLMENT TRENDS
			8	92	ENROLLMENT PROJECTIONS
			9	7843	1-6/OR
			10	240	7-8/OR
8	27	7 AND 6	11	158	10/MAJ
9	27	8/EJ	12	131	11 AND 9
			13	34	12/EJ

STATISTICS ---

	Free Text	Controlled Vocabulary
Total Relevant according to ERIC/HE Assessment = 40 (24)*		
Total Retrieved	27	34
Total Relevant Retrieved	20 (22)	34 (18)
Total Relevant Missed	20 (2)	6 (6)
RECALL	$\frac{20}{40}$ = 50% (92%)	$\frac{34}{40}$ = 85% (75%)
PRECISION	$\frac{20}{27}$ = 74% (81%)	$\frac{34}{34}$ = 100% (53%)

*Some items were assessed differently by the project staff and the Clearinghouse. The project staff assessments, if different, appear in parentheses.

*See ED 160 108

ANALYSIS---

The controlled vocabulary formulation is the better formulation, if you only consider Clearinghouse relevance assessments. Enrollment Trends is not included in the free text formulation. This single element, if incorporated into the free text strategy as part of the prediction facet" (i.e., statements 2 to 5) would retrieve all the citations assessed relevant. Thus, in order to achieve 100% recall, the free text formulation would be restated in the following way:

1 ENROLLMENT?
2 TREND?
3 PREDICT?
4 PROJECTION?
5 FORECAST?
6 FUTUR?
7 1 AND (2 OR 3 OR 4 OR 5 OR 6)
8 .EXECUTE Search Save for HIGHER EDUCATION (see appendix to this report)
9 7 AND 8

The Higher Education Clearinghouse determined that all 34 citations retrieved by the controlled vocabulary formulation are relevant. However, we reassessed the controlled vocabulary output and disagreed with the relevance of 18 citations. In these instances, reference is made to Enrollment Trends, e.g., declining number of students, recruitment practices, student attrition, etc.

An example of the kind of ERIC record which could cause a conflicting relevance judgment follows:

EJ108874 CG508108
Taking the Scare Out of the Student Scarcity
Barton, David W., Jr.
Journal of the National Association of College Admissions Counselors, 19, 2, 4-7 Nov 74
Traces average response to recession, beginning with the use of blinders to maintain the illusion of the good old days, through the belated realization to the ill-considered hysterical response. Then, of course, the business like approach. Presented in Williamsburg, Virginia, in May to the Potomac and Chesapeake ACAC. (Author)
*Descriptors: *Higher Education/ *Business Skills/ *Enrollment Trends/ *Economic Factors/ *Educational Assessment/ State of the Art Reviews/ Organizational Effectiveness*

Some reference or implication about the future has to appear in the record in order for us to decide that it is relevant. Two items judged relevant by the Clearinghouse and us which are not retrieved by the free text search treat current enrollment trends, but seem to suggest the future with index phrases like Educational Demand, Educational Planning, and Labor Market.

The brevity of CIJE abstracts makes this assessment particularly difficult. These two citations, given below, should be incorporated into the ONTAP Manual's answer set for this topic.

EJ116232 HE506383
Current Trends in the United Kingdom
Christopherson, Derman G.
Higher Education, 4, 2, 133-148 May 75
Surveys developments in higher education in the United Kingdom since the publication of the government's 'White Paper' in December 1972. Considers: (1) reorganization of public higher education, (2) extent of student demand, (3) increase in enrollment of women, and (4) financial problems caused by inflation. (Editor/PG)
*Descriptors: *Higher Education/ *Educational Demand/ *Enrollment Trends/ *Females/ Educational Finance/ Financial Problems/ Statewide Planning/ Organizational Change*
*Identifiers: *United Kingdom/ White Paper*

EJ105500 CE501931
The Degree Dilemma
Carbine, Michael E.
Manpower, 6, 11, 2-8 Nov 74
Employer demands for college education rise as enrollments fall. (Author)
*Descriptors: *Higher Education/ *Personnel Policy/ *Enrollment Trends/ *Employment Qualifications/ *Employer Attitudes/ Educational Disadvantagement/ Labor Market/ Subemployment/ Employment Practices/ Employment Level*

Two citations judged non-relevant by HE, relevant by us, were not retrieved by the controlled vocabulary search. They are cited below along with a portion of the text from the record which contributed, in part, to our judgment of relevance.

EJ 107 306 Enrollment Projections (minor descriptor), *Higher Education

EJ 102 812 Enrollment Projections (minor descriptor), *Higher Education, Educational Planning (minor descriptor)

Six citations assessed relevant by both HE and us failed to be retrieved by the controlled vocabulary formulation. They are cited below along with a portion of the text from the record which contributed, in part, to our judgment of relevance.

EJ 114 396 Enrollment Projections (minor descriptor), *Higher Education

EJ 112 284 "and predicts that journalism enrollments will increase. . ." (abstract)

EJ 108 126 *Educational Planning, *Futures (of Society), *Higher Education, Student Enrollment

EJ 104 490 *Higher Education, Enrollment Projections, Educational Planning

EJ 102 836 Challenges to Universities through Year 2000 (title), Enrollment Projections (minor descriptor)

EJ 102 829 Enrollment Projections (minor descriptor), ". . . implications for the future of Higher Education by examining data concerning demographic factors, growth projections . . ." (abstract)

The search topic is very similar to Search Topic #3 which follows. Comparison of these two search strategies points up some interesting differences in both the Controlled Vocabulary and Free Text search strategies. A key difference is the missing concept for trends, and the use of project? for projections.

SEARCH TOPIC #3 (#13 in ERIC ONTAP Manual)**
(not in LRS File 201)

Junior College Enrollment Projections
For the Near Future (1975-95)

(Only RIE items, i.e., ED's)

SEARCH FORMULATION*** ---

In Free Text			In Controlled Vocabulary		
1	832	ENROLLMENT?	1	468	COMMUNITY COLLEGES
2	4810	PROJECT?	2	841	JUNIOR COLLEGES
3	1750	FUTUR?	3	92	ENROLLMENT PROJECTIONS
4	462	ESTIMAT?	4	22	(1 OR 2) AND 3
5	90	FORECAST?	5	21	4/ED
6	1820	.EXECUTE Search Save for JUNIOR COLLEGES (see Appendix)			
7	73	1 AND 6 AND (2 OR 3 OR 4 OR 5)			
8	68	7/ED			

STATISTICS ---

	Free Text	Controlled Vocabulary
Total Relevant according to ERIC/JC Assessment = 19 (24)*		
Total Retrieved	68	21
Total Relevant Retrieved	19 (24)	15
Total Relevant Missed	0	6 (10)
RECALL	$\frac{19}{19}$ = 100%	$\frac{15}{19}$ = 79% (66%)
PRECISION	$\frac{19}{68}$ = 28% (35%)	$\frac{15}{21}$ = 71%

*Some items were assessed differently by the project staff and the Clearinghouse. The project staff assessments, if different, appear in parentheses.

**See ED 160 108

***The free text search is a composite search developed for high recall (see footnote on page 1 of this report.) An ERIC Clearinghouse staff person contributed the controlled vocabulary search.

ANALYSIS ---

Neither search formulation is better. The free text formulation results in low precision, but higher recall.

There are conflicting relevance assessments for seven of the 68 retrieved items retrieved by the free text strategy. Difficulty is encountered when determining the relevance of articles which refer to the future (e.g., student population and estimates of enrollments), but on the whole, describe the present. On one hand, the article treats present trends; on the other, it makes implications for the future. It may be argued that items containing enrollment projections for postsecondary education may take into account community college enrollment projections since postsecondary education deals with all education beyond the secondary level. A summary of conflicting relevance assessments is given below:

Number of Junior College assessed relevant items judged non-relevant by us	19 0
Number of Junior College assessed non-relevant items judged relevant by us reassessed non-relevant by us	5 2

The content of the six non-relevant citations which were retrieved by both formulations can be divided into three types:

1. articles which present methods and techniques for making enrollment projections (ED 110 101, ED 103 070, ED 096 540);
2. those treating enrollment projections for a single institution (ED 104 484); or
3. for a time period other than 1975-95 (ED 101 221, ED 101 650).

Many of the 43 non-relevant citations culled by the free text strategy fall into one of these three groups. The remainder deals with enrollment projections for postsecondary education rather than Junior Colleges specifically.

Leaving out the search term and its variants, Post Secondary, a reformulation of the search save for Junior Colleges is suggested in order to eliminate some of the non-relevant material from the output. Excluding these search terms from the free text formulation results in a low recall since some relevant items contain references to postsecondary but not Junior College Education.

The controlled vocabulary formulation achieves high precision, but fails to retrieve four relevant items, according to the Junior College Clearinghouse assessment. Portions of text from these four records are cited below in an effort to show why they are not retrieved.

ED 105 958 A Perspective for the Future, Community Colleges for the 1980's (title); *Futures (of Society) (descriptor); "Includes charts illustrating the growth and decline of enrollments (1880-2000) . . ." (abstract)

ED 099 062 *Enrollment Trends, "Community Colleges (descriptor); "Information was obtained . . . in order to examine the validity of earlier forecasts of manpower requirements, population, and enrollments, and to enable future projections . . ." (abstract)

ED 098 881 ". . . to accommodate the enrollments between 1975 and 1990 . . . emphasis is placed on projections of postsecondary enrollments . . ." (abstract). (Although no mention is made of Junior Colleges, it may be reasoned that this article deals with Junior Colleges as part of secondary education.)

ED 095 782 ". . . has presented population trends that influence postsecondary enrollments to 1984. Concerning these projections . . ." (abstract). (Like ED 098 881, this deals with the area of postsecondary education.)

The search is similar to Search Topic #2. A comparison of the two points up the value of the missing descriptor in the Controlled Vocabulary search: Enrollment Trends, the need for the free text word trends, and the weakness in the use of the truncation for projection(s), i.e., Project?.

SEARCH TOPIC #4 (#8 in ERIC ONTAP Manual)**

Drug Abuse, Including Alcohol, Among
Students of Elementary and Secondary Schools

SEARCH FORMULATIONS --- The free text search is a composite search developed for high recall (see footnote on page 1 of this report.) An ERIC Clearinghouse staff person contributed the controlled vocabulary search.

In Free Text			In Controlled Vocabulary		
1	6903	.EXECUTE Search Save for ELEMENTARY EDUCATION (see Appendix)	1	152	DRUG ABUSE
2	7644	.EXECUTE Search Save for SECONDARY EDUCATION (see Appendix)	2	33	DRUG ADDICTION
3	128	ALCOHOL?	3	7	LYSERGIC ACID DIETHYLAMIDE
4	178	DRUG(W)ABUSE	4	31	MARIHUANA
5	34	DRUG(W)ADDICTION	5	22	NARCOTICS/DF
6	3	LSD	6	35	DRINKING/DF
7	7	LYSERGIC	7	23	ALCOHOLIC BEVERAGES
8	311	3-7/OR	8	56	ALCOHOLISM/DF
9	76	1 AND 2 AND 8	9	254	1-8/OR
			10	6284	.EXECUTE Search Save for ELEMENTARY EDUCATION (controlled vocabulary only)(see appendix to this report)
			11	7020	.EXECUTE Search Save for SECONDARY EDUCATION (controlled vocabulary only)(see appendix to this report)
			12	44	9-11/AND

STATISTICS ---

	Free Text	Controlled Vocabulary
Total Relevant according to ERIC/SP = 64 (66)*		
Total Retrieved	76	44
Total Relevant Retrieved	63 (65)	41 (42)
Total Relevant Missed	1	23 (24)
RECALL	$\frac{63}{64}$ = 98%	$\frac{41}{64}$ = 64% (63%)
PRECISION	$\frac{63}{76}$ = 82% (85%)	$\frac{41}{44}$ = 93% (95%)

*Some items were assessed differently by the project staff and the Clearinghouse. The project staff assessments, if different, appear in parentheses.

** See ED 160 108

ANALYSIS ---

The free text search results are better. In the reformulated controlled vocabulary strategy, the searcher intentionally omitted descriptors pertinent to the area of Drug Education, e.g., Drug Education, Health Education, Alcohol Education. Consequently, 16 relevant citations which deal with drug abuse in the schools but are assigned drug education descriptors fail to be retrieved by the controlled vocabulary strategy.

Three others were missed because the relevant citations contain information in the abstract on student drug abuse, but the indexing record does not include grade level descriptors.

Two items are provided below in which drug abuse among school students is discussed but not within the school setting. Population descriptors, e.g., adolescents, youth, and grade level index terms may be searched to find all relevant items.

ED103756 CG009627
A Strategic Study for Communication Programs on Alcohol and Highway Safety.
National Highway Traffic Safety Administration (DOT). Washington, D.C.
Nov 74 76p.
EDRS Price MF-$0.76 HC-$4.43 PLUS POSTAGE
This report is a synopsis of the findings of a nine-month research study into the beliefs, attitudes, and behavior of both adults and high school youth involved in the use of alcoholic beverages and driving. A questionnaire was administered to a nationally projectable sample of 1,660 adults between the ages of 18 and 55 and to a representative sample of 397 high school students. The questionnaire attempted to measure the success of an extensive three-year public information/communications effort on the problem of drunk driving, sponsored by the U.S. Department of Transportation's National Highway Traffic Safety Administration (NHTSA). Results indicated significant progress since 1970 in increasing people's awareness of the drunk driving problem, as well as a readiness to pay higher taxes to support stricter law enforcement programs. The study among high school students revealed an alarming amount of drinking and driving, and almost as much involvement with alcohol-related situations as adults. The report also provides a thorough evaluation of both sample groups which can be used to give direction to future efforts to solve the drunk driving problem. (Author/PC)
*Descriptors: Adolescents/ Adults/ Alcoholic Beverages/ *Attitudes/ *Beahvior Patterns/ *Drinking/ Research Projects/ *Social Problems/ Surveys/ *Traffic Safety*

EJ110132 UD503323
Inter- and Intragenerational Influences on Adolescent Marijuana Use
Kandel, Denise
Journal of Social Issues, 30, 2, 107-136 Spr 74
A survey of a representative sample of high school students in New York State revealed that involvement with other drug using adolescents is a more important correlate of marijuana use than is parental use of psychoactive drugs or alcohol. Implications of these data for parental and peer influence, the generation gap, and social change are discussed. (Author/EH)
*Descriptors: *Marijuana/ *Peer Relationship/ *Parent Child Relationship/ *Generation Gap/ *Parent Influence/ Peer Groups/ Social Change/ Youth Problems*

The remaining relevant citations, missed by the controlled vocabulary formulation, mention drug abuse in the abstract (e.g., as a chapter in a curriculum on science education, or as one topic in a set of guidelines for school policy). In these instances, drug abuse, comprising one of many topics in a multi-faceted article, is not assigned a descriptor.

The controlled vocabulary formulation retrieves one relevant item on alcohol abuse which is missed by the free text strategy. In order to include this item in the free text output, the strategy should be revised to include the search term Drinking. However, this results in culling three additional non-relevant documents on secondary school science and drinking water. The examples which follow point out false drops retrieved when Drinking is employed as a free text term:

EJ116612 SE513475
WHAT DO YOU HAVE FOR OPENERS?
LLEWELLYN, GERALD C.
SCIENCE ACTIVITIES, 12, 2, 10-12 MAR/APR 75
DESCRIBES SOME "OPENERS" TO USE IN SCIENCE CLASS TO ATTRACT ATTENTION, CREATE INTEREST OR MAINTAIN INTEREST. SOME EXAMPLES OF MATERIALS THAT CAN BE USED ARE A RADIOMETER, A [DRINKING] BIRD, GEIGER-MULLER COUNTER AND A HUMAN SKELETON. (BR)
*DESCRIPTORS: *INSTRUCTIONAL MATERIALS/ *MOTIVATION/ *SCIENCE ACTIVITIES/ *SCIENCE EQUIPMENT/ INSTRUCTION/ SCIENCE EDUCATION/ SECONDARY EDUCATION/ SECONDARY SCHOOL SCIENCE/ THREE DIMENSIONAL AIDS*

EJ113272 SE512820
[DRINKING] WATER--HOW ARE WE COPING?
HAWKINS, MARY E.
SCIENCE TEACHER, 42, 2, 16-20 FEB 75

To show how this search was conceptualized using a building block approach in free text, we have excerpted from the ONTAP Manual:

* Step 6: Conceptualizing the Search as Input to the Retrieval System

Each search statement in the Briefsearch given below is an example of the building block approach as both facets, DRUG ABUSE and K-12 STUDENTS are present. Search terms chosen for the Briefsearch are derived from ERIC descriptors having high postings or used frequently in descriptor phrases.

Briefsearch	Comments
1 S ALCOHOL(C)SCHOOL(C)STUDENTS	The (C) limiter is used rather than the AND in order to reduce the number of individual SELECT and COMBINE operations
2 S DRUG(W)ABUSE(C)HIGH(W)SCHOOL	
3 S DRUG(W)ABUSE(C)ELEMENTARY	
4 C 1-3/OR	

A high precision formulation relying solely on the descriptor phrase Drug Abuse as a descriptor, identifier, or title word, to create the DRUG ABUSE facet and descriptors for grade levels is shown below:

Formulation #2

1 S DRUG(W)ABUSE/DE,ID,TI

2 S ELEMENTARY SCHOOL STUDENTS

3 S HIGH SCHOOL STUDENTS

4 S SECONDARY SCHOOL STUDENTS

5 C (2 OR 3 OR 4) AND 1

Comprehensive Search Saves for grade levels K-8 (Search Save #2, see appendix) and grade levels 9-12 (Search Save #3, see appendix) are used in the following high recall formulation to create the K-12 STUDENTS facet. The DRUG ABUSE facet is comprised of individual drug names in addition to general terms such as narcotics, drug addiction, etc.

*From Markey, Karen and Pauline Atherton. ONTAP: Online Training and Practice Manual for ERIC Data Base Searchers. (Syracuse, NY: ERIC Clearinghouse on Information Resources, Syracuse University, June 1978), pp. 101-103. ED 160 108

Formulation #3		Comments
1	.EXECUTE_____ (Search Save for ELEMENTARY EDUCATION)	Statements 1 and 2 create the set for K-12 STUDENTS
2	.EXECUTE_____ (Search Save for SECONDARY EDUCATION)	
3	S ALCOHOL?	Included DESCRIPTORS and Identifiers: ALCOHOL EDUCATION, ALCOHOLIC BEVERAGES, ALCOHOLISM, Alcohol, 4 multi-term alcohol identifiers, DRUG ABUSE, 5 multi-term drug abuse identifiers, DRUG ADDICTION, LYSERGIC ACID DIETHYLAMIDE, MARIHUANA, NARCOTICS, 2 multi-term Narcotics identifiers
4	S DRUG(W)ABUSE	
5	S DRUG(W)ADDICTION	
6	S LSD	
7	S LYSERGIC	
8	S MARI?UANA?	DIALOG's truncation feature allows variant spellings of the term to be retrieved
9	S NARCOTIC?	
10	C 3-9/OR	
11	C 10 AND (1 OR 2)	

SEARCH TOPIC #5 (#5 in ERIC ONTAP Manual)**

Use of School Busing to Achieve Racial Integration

SEARCH FORMULATIONS -- The free text search is a composite search developed for high recall (see footnote on page 1 of this report.) An ERIC Clearinghouse staff person contributed the controlled vocabulary search.

	In Free Text			In Controlled Vocabulary		Comments
1	48	BUS	1	41	STUDENT TRANSPORTATION	Using truncation to retrieve the many possible forms of BUS would probably result in a large number of false drops. Each form is typed in separately rather than entering BUS?
2	1	BUSSES	2	27	BUS TRANSPORTATION	
3	14	BUSES	3	8	SCHOOL BUSES	
4	2	BUSSING	4	62	1-3/OR	
5	33	BUSING	5	604	R1-R15 (online use of Thesaurus descriptor display) R1: RACIAL SEGREGATION	
6	3	BUSED				
7	1	BUSSED				
8	98	STUDENT(C)TRANSPORTATION	6	556	R1, R6-R7, R10-R11, R15-R21, R24, R26, R27-R30 R1: RACIAL INTEGRATION	Included DESCRIPTORS/Identifiers: BUS TRANSPORTATION, SCHOOL BUSES/Bus Drivers, Bus Driver Training.
9	155	1-8/OR				
10	800	RACIAL?				
11	1306	INTEGRAT?	7	539	R1-R7, R10, R13-R17 R1: RACE RELATIONS	Included DESCRIPTORS/Identifiers: 22 INTEGRATED & INTEGRATION DESCRIPTORS, SEGREGATED PUBLIC FACILITIES, 7 SEGREGATION DESCRIPTORS, SEGREGATIONIST ORGANIZATIONS/ Desegregation Aid, Desegregation Advisory Project.
12	127	DESEGREGAT?				
13	151	SEGREGAT?	8	810	5-7/OR	
14	1947	10-13/OR	9	29	7 AND 8	
15	63	9 AND 14	10	5	BUSING/ID	
			11	34	9 OR 10	

STATISTICS ----

	Free Text	Controlled Vocabulary
Total Relevant according to ERIC/SO = 46 (45)*		
Total Retrieved	63	34
Total Relevant Retrieved	46 (45)	34 (33)
Total Relevant Missed	0	12
RECALL	$\frac{46}{46}$ = 100%	$\frac{34}{46}$ = 74% (73%)
PRECISION	$\frac{46}{63}$ = 73% (71%)	$\frac{34}{34}$ = 100% (97%)

*Some items were assessed differently by the project staff and the Clearinghouse. The project staff assessments, if different, appear in parentheses.

**See ED 160 108

ANALYSIS ---

The comparison of the search results makes one see how valuable the search objective is for determining the optimum formulation. The three descriptors used in the controlled vocabulary formulation to represent the facet of Busing were not sufficient for retrieving all the relevant items. Only five of the 46 relevant citations were assigned Busing as an identifier. The term busing appears in the abstract and/or title of all but one of the 12 relevant documents missed by the controlled vocabulary strategy. School Integration is assigned as a descriptor to half of these missed items, but no other descriptor implying Busing is common to these 12 citations. Portions of text relevant to the topic and their locations in several missed items are cited below:

EJ 115 853 "And anti-busing legislation" (title)

EJ 115 849 ". . . to the present day strife over bussing." (abstract)

EJ 113 526 ". . . to use cross-busing in order to remedy racial segregation" (abstract)

EJ 107 317 "racial integration and school busing in Boston" (abstract)

EJ 110 551 "Nixon's proposed anti-busing program constitutes the desegregation area." (abstract)

EJ 103 288 ". . . examines the anti-busing movement in the North." (abstract)

The question of which is better? depends completely on the search objective: if it is high recall, then free text is better; if it is high precision, then controlled vocabulary is better.

SEARCH TOPIC #6 (#10 in ERIC ONTAP Manual)*

Jean Piaget's Theories, and Thought Processes or Language Development of Children

SEARCH FORMULATIONS --- The free text search is a composite search developed for high recall (see footnote on page 1 of this report.) An ERIC Clearinghouse staff person contributed the controlled vocabulary search.

In Free Text		
1	238	PIAGET?
2	93	1/AVAIL

In Controlled Vocabulary		
1	238	PIAGET?
2	433	COGNITIVE DEVELOPMENT
3	96	INTELLECTUAL DEVELOPMENT
4	391	COGNITIVE PROCESSES
5	150	THOUGHT PROCESSES
6	35	ABSTRACT REASONING
7	324	CONCEPT FORMATION
8	53	ABSTRACTION LEVELS
9	44	CREATIVE THINKING
10	157	CRITICAL THINKING
11	32	GENERALIZATION/DF
12	99	LOGICAL THINKING
13	25	PRODUCTIVE THINKING
14	596	PROBLEM SOLVING
15	260	LEARNING THEORIES
16	197	EDUCATIONAL THEORIES
17	33	VERBAL DEVELOPMENT
18	462	LANGUAGE DEVELOPMENT
19	61	VERBAL ABILITY
20	194	VERBAL COMMUNICATION
21	79	LINGUISTIC COMPETENCE
22	261	PSYCHOLINGUISTICS
23	177	ORAL COMMUNICATION
24	74	ORAL EXPRESSION
25	3097	2-24/OR
26	182	1 AND 25
27	77	26/AVAIL

STATISTICS ---

	Free Text	Controlled Vocabulary
Total Relevant = 87		
Total Retrieved	93	77
Total Relevant Retrieved	87	76
Total Relevant Missed	0	11
RECALL	$\frac{87}{87}$ = 100%	$\frac{76}{87}$ = 87%
PRECISION	$\frac{87}{93}$ = 94%	$\frac{76}{77}$ = 99%

*See ED 160 108

ANALYSIS ---

The free text search was the better of the two. The controlled vocabulary formulation submitted by the Early Childhood Clearinghouse is the only one of six discussed which is partially comprised of free text, i.e., search statement #1: PIAGET?. Imposing some 20 descriptors on the results of this statement reduces the output from 93 to 77 citations, of which all but one are relevant.

Concerning the 11 relevant citations missed by the controlled vocabulary search, eight listed Piaget, Jean as an identifier; six of the eight could have been retrieved by including Child Development and Developmental Psychology in the list of descriptors on Piagetian theories. The five citations which would still be outstanding are as follows, accompanied by text relevant to the topic:

ED 106 721 ". . . the concepts of Piaget and the application to classroom learning problems of Piaget's stages in intellectual development" (abstract)

ED 100 510 *Piaget (Jean) (identifier); "reviews infant learning literature . . . and discusses Piaget's theory of development" (abstract)

ED 104 530 ". . . approach to child development incorporating theories of Piaget . . ." (abstract)

ED 098 972 Piaget (Jean) (identifier); "This approach, drawing frequently on Piagetian theory . . . provides . . . picture of the learner and his changing cognitive structures." (abstract)

ED 107 345 *Piaget (Jean) (identifier); *Educational Psychology, *Behavioral Theories (descriptor)

As the free text searcher experimented with several search strategies, we thought it best to show the reasoning behind the final choice based on several considerations and the differing results. This is another excerpt from the ONTAP Manual:

*Step 6: Conceptualizing the Search as Input to the Retrieval System

A Briefsearch, performced initially before a complex formulation, gives a rough estimation of the size of the final output. Records from Briefsearch results can be reviewed in order to collect additional searching vocabulary.

Briefsearch

1 S PIAGET(C)COGNITIVE

2 S PIAGET(C)CHILDREN

3 C 1 OR 2

4 LIMIT 3/AVAIL

At this point 72 citations are retrieved. Searching on the most specific facet, PIAGET, and comparing the results of this statement with that of the Briefsearch gives us a general idea how large the final set may be.

Formulation #2	Comments
1 S PIAGET?	Included DESCRIPTORS and identifiers: Piaget 17 multi-term Piaget identifiers, Piagetian, 9 multi-term Piagetian identifiers.
2 LIMIT 1/AVAIL	This search retrieves 93 citations. Since the requester wants all relevant items, the searcher may consider terminating the search at this point, having determined that the volume of output is small enough for manual review.

The addition of the facet CHILDREN to the formulation reduces the output to 77 citations. This size of output may be considered sufficient to forego the introduction of another facet to the formulation.

Formulation #3	Comments
3 . EXECUTE _____(Search Save for CHILDREN)	Search Save #4 (see appendix) is used for the CHILDREN facet
4 C 2 AND 3	

The fourth formulation is comprised of two facets, THEORIES and PIAGET. It requires much typing so that using the online thesaurus may be favored instead of employing free text searching.

---

*From Markey, Karen and Pauline Atherton. ONTAP: Online Training and Practice Manual for ERIC Data Base Searchers. (Syracuse, NY: ERIC Clearinghouse on Information Resources, Syracuse University, June 1978), pp. 112-116. ED 160 108

Formulation #4		Comments
1	S PIAGET?	Just like in formulation #3, the first two statements are taken from formulation #2 and make up the entire facet PIAGET.
2	LIMIT 1/AVAIL	
3	S AFFECTIVE(W)BEHAVIOR	Included DESCRIPTORS/Identifiers: AFFECTIVE BEHAVIOR
4	S COGNITIVE(W)ABILITY	COGNITIVE ABILITY
5	S COGNITIVE(W)ABILITIES	Cognitive Abilities Test
6	S COGNITIVE(W)DEVELOPMENT	COGNITIVE DEVELOPMENT/Cognitive Development Research Tools
7	S COGNITIVE(W)DEVELOPMENTS	
8	S COGNITIVE(W)OBJECTIVE	
9	S COGNITIVE(W)OBJECTIVES	COGNITIVE OBJECTIVES
10	S INTELLECTUAL(W)DEVELOPMENT	INTELLECTUAL DEVELOPMENT
11	S INTELLECTUAL(W)DEVELOPMENTS	
12	S MENTAL(W)DEVELOPMENT	MENTAL DEVELOPMENT
13	S PERCEPTUAL(W)DEVELOPMENT	PERCEPTUAL DEVELOPMENT
14	S ECHOLALIA	ECHOLALIA
15	S FLES	FLES, FLES GUIDES, FLES MATERIALS, FLES PROGRAMS, FLES TEACHERS/FLES Objectives, FLES Program
16	S READING(W)HABIT	
17	S READING(W)HABITS	READING HABITS
18	S HUMAN(W)DEVELOPMENT	HUMAN DEVELOPMENT/3 multi-term Human Development Identifiers
19	S LANGUAGE?	LANGUAGE, over 60 multi-term LANGUAGE DESCRIPTORS, 31 LANGUAGES DESCRIPTORS/ 50 multi-term Language Identifiers
20	S NUCLEATION	NUCLEATION (LANGUAGE LEARNING)
21	S PSYCHOLINGUISTIC?	PSYCHOLINGUISTICS
22	S READING(W)DEVELOPMENT	READING DEVELOPMENT
23	S LINGUISTIC?	LINGUISTICS, 13 multi-term LINGUISTICS and LINGUISTIC DESCRIPTORS/26 multi-term Linguistic and Linguistics Identifiers
24	S SPEECH(W)DEVELOPMENT	RETARDED SPEECH DEVELOPMENT
25	S SPEECH(W)HABITS	SPEECH HABITS
26	S VERBAL(W)DEVELOPMENT	VERBAL DEVELOPMENT
27	S VOCABULARY(W)DEVELOPMENT	VOCABULARY DEVELOPMENT
28	S HANDWRITING(W)DEVELOPMENT	HANDWRITING DEVELOPMENT
29	C 3-28/OR	Cumulative set for cognitive and language development
30	C 29 AND 2	

Imposing the facet PIAGETIAN THEORIES reduces the final set to 71 citations.

At this point, the searcher terminates the search in order to avoid losing rele-

vant material. If a decision is made to include the CHILDREN facet, use could be made of the Search Save demonstrated in formulation #3.

30 .EXECUTE _____ (Search Save for CHILDREN)

31 C 30 AND 29

After the CHILDREN facet is incorporated, the total number of retrieved citations is 60.

(Reader: In light of the requestor's desire to retrieve all relevant citations regardless of cost, which strategy do you prefer? _____)

Step 8: Evaluating Final Results

Formulation #2 in which the most specific facet of the topic is searched retrieves all 81 relevant citations. It is the only formulation of those given which achieves 100% recall.

Formulation #2

Number of citations found = 93

$$\text{Recall} = \frac{\text{Number of relevant citations found}}{\text{Number of citations in answer set}} \quad \frac{81}{81}$$ (100% relevant documents retrieved)

$$\text{Precision} = \frac{\text{Number of relevant citations found}}{\text{Number of citations found}} \quad \frac{81}{93}$$ (87% of retrieved documents are relevant

The Briefsearch achieves high precision and high recall but does not retrieve all items relevant to the topic. But, used as a brief survey of the file, it aids in determining whether to impose a second facet to the formulation.

Briefsearch

Number of citations found = 72

$$\text{Recall} = \frac{\text{Number of relevant citations found}}{\text{Number of citations in answer set}} \quad \frac{66}{81}$$ (81% relevant documents retrieved)

$$\text{Precision} = \frac{\text{Number of relevant citations found}}{\text{Number of citations found}} \quad \frac{66}{72}$$ (92% of retrieved documents are relevant)

Two facets, CHILDREN and PIAGET, of the topic make up formulation #3. Imposing the CHILDREN aspect into the formulation eliminates 14 relevant documents from the output.

Formulation #3

Number of citations found = 77

Recall = $\frac{\text{Number of relevant citations found}}{\text{Number of citations in answer set}}$ $\frac{67}{81}$ (83% relevant documents retrieved)

Precision = $\frac{\text{Number of relevant citations found}}{\text{Number of citations found}}$ $\frac{67}{77}$ (87% of retrieved documents are relevant)

All facets of the topic are included in the building block approach, formulation #4. If the search is terminated before the CHILDREN facet is included in the formulation, 71 citations are retrieved; after the imposition of the CHILDREN facet, 60 citations are retrieved.

Formulation #4. ("Before" CHILDREN facet, statements 1-31)

Number of citations found = 71

Recall = $\frac{\text{Number of relevant citations found}}{\text{Number of citations in answer set}}$ $\frac{71}{81}$ (88% relevant documents retrieved)

Precision = $\frac{\text{Number of relevant citations found}}{\text{Number of citations found}}$ $\frac{71}{71}$ (100% of retrieved documents are relevant)

Formulation #4. ("After" CHILDREN facet, statements 1-33)

Number of citations found = 60

Recall = $\frac{\text{Number of relevant citations found}}{\text{Number of citations in answer set}}$ $\frac{60}{81}$ (75% relevant documents retrieved)

Precision = $\frac{\text{Number of relevant citations found}}{\text{Number of citations found}}$ $\frac{60}{60}$ (100% of retrieved documents are relevant)

As more facets are added to the formulation in the third and fourth examples, the recall ratio decreases. Since high recall is one of the search objectives, formulation #2 is preferred. Not only does it achieve a high recall score, but it is efficient and rapid in terms of online connect time and pre-search preparation.

OBSERVATIONS AND CONCLUSIONS

(A) Relevance Assessments

As a result of the Clearinghouse review of output and controlled vocabulary formulations, there exists a discrepancy concerning the relevance of 11% of the total output for these six searches (392 items). We reassessed the original relevance decisions performed by Charles Bourne at ILR and our subsequent work. In one out of four cases, we reversed an earlier decision. Unresolved in respect to relevance is 8% of the obtained output, or 30 of 392 items.

(B) Recall/Precision Performance

Table 2 summarizes precision and recall measures according to Clearinghouse and our assessment for both controlled vocabulary and free text searches.

Table 2

Precision and Recall of Online Searching (Summary)

	Recall	%	Precision	%
Controlled Vocabulary Formulation				
Clearinghouse assessment	$\frac{215}{283}$	76	$\frac{215}{226}$	95
Our assessment	$\frac{199}{273}$	(72)	$\frac{199}{225}$	(88)
Free Text Formulation				
Clearinghouse assessment	$\frac{261}{282}$	93	$\frac{270}{372}$	71
Our assessment	$\frac{270}{273}$	(99)	$\frac{270}{372}$	(73)

Table 3 gives these data in more detail, search by search.

Table 3

Recall and Precision for Six Searches

(Based on CH assessments)

	Recall	%	Precision	%
Controlled Vocabulary Formulation				
1. Exceptional Children	$\frac{15}{27}$	56	$\frac{15}{17}$	88
2. Higher Education	$\frac{34}{40}$	85	$\frac{34}{34}$	100
3. Junior Colleges	$\frac{15}{19}$	79	$\frac{15}{21}$	71
4. Teacher Education	$\frac{41}{64}$	64	$\frac{41}{44}$	93
5. Social Studies	$\frac{34}{46}$	74	$\frac{34}{34}$	100
6. Early Childhood	$\frac{76}{87}$	87	$\frac{76}{77}$	99
Free Text Formulation				
1. Exceptional Children	$\frac{27}{27}$	100	$\frac{27}{45}$	60
2. Higher Education	$\frac{20}{40}$	50	$\frac{20}{27}$	74
3. Junior Colleges	$\frac{19}{19}$	100	$\frac{19}{68}$	28
4. Teacher Education	$\frac{63}{64}$	98	$\frac{63}{76}$	82
5. Social Studies	$\frac{46}{46}$	100	$\frac{46}{63}$	73
6. Early Childhood	$\frac{87}{87}$	100	$\frac{87}{93}$	94

These figures show that:

1. The controlled vocabulary searches achieve higher precision than the free text searches.

2. Likewise, the free text searches result in higher recall than the controlled vocabulary searches, except in Search #2.

3. Although varying from 28% to 87% on two separate search topics, precision for free text searching is 71%. Only three out of ten citations from the output of the six free text formulations are non-relevant.

4. Clearinghouse personnel, formulating the search topics in controlled vocabulary, retrieved relevant items not found by the free text formulation.

Some important points which are suggested by the Tables above:

1. Recall of 50% for the Higher Education free text search is remarkably low compared to the other five scores, and in fact, is the lowest recall score of all reported. It is important to note that conflicting relevance assessments make up 18 of the 40 relevant citations for this search.

2. Precision for the free text searches varies from 28% to 87%. However, four of the six free text precision scores are higher than the lowest precision score (71%) for the Junior College controlled vocabulary search.

(C) Overlap

From the following Table it is clear that two out of every three citations are retrieved by both controlled vocabulary and free text searches.

Table 4

Overlap of Items Retrieved
by Controlled and Free Text Searches

	Topics	Total Relevant (est.)	Relevant Items Retrieved by Both Controlled and Free Text Searches	% Overlap
1.	Exceptional Children	27	15	56
2.	Higher Education	40	14	35
3.	Junior Colleges	19	15	79
4.	Teacher Education	64	40	63
5.	Social Studies	46	34	74
6.	Early Childhood	87	76	87

Non-matched items account for 21 citations, 18 of which are conflicting relevance assessments; the three remaining relevant items failed to be retrieved by the free text searches.

The anaiysis of controlled vocabulary and free text searches demonstrates that:

1. Searching cannot be limited to controlled vocabulary fields in the ERIC data base if the search objective is to achieve 100% recall.

2. Free text did not retrieve all the relevant citations, but achieved 93% recall, while controlled vocabulary searches only retrieved 76% of the relevant citations.

3. Precision was not so low for free text searching as to be burdensome. Seven of ten citations retrieved by free text were relevant.

4. Free text retrieved 261 relevant items; controlled vocabulary retrieved 215 relevant citations; 194 relevant citations were retrieved by both formulations, i.e., 66% of the controlled vocabulary output duplicated the free text output.

RECOMMENDATIONS

Free text searching can often be the better search formulation. This fact alone should influence data base design, for the searchable fields should be more consistent in word form. Replacing cryptic, non-uniform text/terms should be possible when ERIC's basic index is reviewed and revised for inconsistencies. (See Part IV of this project report for more on this.)

Experienced users of ERIC online should prefer free text searching whenever high recall is a search objective.

This test should be enlarged to include all DIALOG ONTAP ERIC sample search questions. See Appendix.

APPENDIX

In our tests, we used some Save Searches for facets of searches not easily represented by a single descriptor:

(1) Higher Education

(2) Junior College

(3) Elementary Education

(4) Secondary Education

These are included here for easy reference.

Also included here is DIALOG's complete ONTAP ERIC Sample Search Question List (file 201).

HIGHER EDUCATION

Search Save for: Higher Education

The terms COLLEGE, COLLEGES, HIGHER EDUCATION, GRADUATE SCHOOL, UNIVERSITY, and UNIVERSITIES are searchable in all subject-conveying fields except corporate source, sponsoring agency and descriptive note fields. This reduces the number of false drops which may occur when a particular educational institution is responsible for the document as a corporate author or sponsor.

FORMULATION	COMMENTS AND DESCRIPTORS/Identifiers
1 S COLLEGE/DE,ID,TI,AB	COLLEGE ADMINISTRATION, 40 other multi-term COLLEGE Descriptors/College, over 60 multi-term College Identifiers
2 S COLLEGES/DE,ID,TI,AB	COLLEGES, 13 multi-term COLLEGES Descriptors
3 S COLLEGIA?	Collegial System, Collegial Teams, four multi-term Collegiate Identifiers
4 S DOCTORAL	DOCTORAL DEGREES, POST DOCTORAL EDUCATION, DOCTORAL PROGRAMS, DOCTORAL THESES/ Doctoral These, Doctoral Thesis
5 S GRADUATE(W)SCHOOL/DE,ID,TI,AB	Graduate School Foreign Language Test, Graduate School of Education
6 S GRADUATE(W)SCHOOLS	Graduate Schools, 2 multi-term Graduate Schools Identifiers
7 S GRADUATE(W)STUDENT	
8 S GRADUATE(W)STUDENTS	GRADUATE STUDENTS
9 S GRADUATE(W)STUDY	GRADUATE STUDY
10 S GRADUATE(W)STUDIES	
11 S HIGHER(W)EDUCATION/DE,ID,TI,AB	HIGHER EDUCATION/32 multi-term Higher Education Identifiers
12 S POST(W)SECONDARY	POST SECONDARY EDUCATION/Post Secondary Schools
13 S POSTSECONDARY	
14 S UNDERGRADUATE?	UNDERGRADUATE STUDENTS, UNDERGRADUATE STUDY/ 5 multi-term Undergraduate Identifiers

Search Save for: (cont.)

FORMULATION	COMMENTS AND DESCRIPTORS/Identifiers
15 S UNIVERSITY/DE,ID,TI,AB	Three multi-term UNIVERSITY Descriptors/ University, over 100 multi-term Identifiers
16 S UNIVERSITIES/DE,ID,TI,AB	UNIVERSITIES, 3 multi-term UNIVERSITIES Descriptors/3 multi-term Universities Identifiers
17 C 1-16/OR	

Approximate Processing Time: .076 hrs. ($1.90) in file 1.

JUNIOR COLLEGE

Search Save for: Junior College

The terms JUNIOR COLLEGE and COMMUNITY COLLEGE are searchable in all subject-conveying fields except corporate source, sponsoring agency and descriptive note in order to exclude items sponsored and/or authored by a specific educational institution.

FORMULATION	COMMENTS AND DESCRIPTORS/Identifiers
1 S ASSOCIATE(W)DEGREE	
2 S ASSOCIATE(W)DEGREES	ASSOCIATE DEGREES
3 S COMMUNITY(W)COLLEGE/DE, ID,TI,AB	8 multi-term Community College Identifiers
4 S COMMUNITY(W)COLLEGES	COMMUNITY COLLEGES
5 S GRADE 13	GRADE 13
6 S GRADE 14	GRADE 14
7 S JUNIOR(W)COLLEGE/DE,ID, TI,AB	JUNIOR COLLEGE LIBRARIES, JUNIOR COLLEGE STUDENTS/6 multi-term Junior College Identifiers
8 S JUNIOR(W)COLLEGES	JUNIOR COLLEGES
9 S POST(W)SECONDARY	POST SECONDARY EDUCATION/Post Secondary Schools
10 S POSTSECONDARY	
11 C 1-10/OR	

Approximate Processing Time: .049 hrs. ($1.23) in file 1.

ELEMENTARY EDUCATION

Search Save for: Elementary Education (grades K-8)

FORMULATION	COMMENTS AND DESCRIPTORS/Identifiers
1 S EARLY(W)CHILDHOOD	EARLY CHILDHOOD, EARLY CHILDHOOD EDUCATION, 14 multi-term Early Childhood identifiers
2 S ELEMENTARY(1W)EDUCATION	ELEMENTARY EDUCATION, ELEMENTARY SECONDARY EDUCATION/Elementary Education Voucher Demonstration, 16 multi-term variations of Elementary Secondary Education Act
3 S ELEMENTARY(W)GRADE	
4 S ELEMENTARY(W)GRADES	ELEMENTARY GRADES
5 S ELEMENTARY(W)PROGRAM	
6 S ELEMENTARY(W)PROGRAMS	UNGRADED ELEMENTARY PROGRAMS/Elementary Programs
7 S ELEMENTARY(W)SCHOOL	ELEMENTARY SCHOOL COUNSELING, E.S. COUNSELORS, E.S. CURRICULUM, E.S. GUIDANCE, E.S. LIBRARIES, E.S. MATHEMATICS, E.S. ROLE, E.S. SCIENCE, E.S. STUDENTS, E.S. SUPERVISORS, E.S. TEACHERS/ Elementary School, 11 multi-term Elementary School identifiers
8 S ELEMENTARY(W)SCHOOLS	ELEMENTARY SCHOOLS, BIRACIAL ELEMENTARY SCHOOLS, CATHOLIC ELEMENTARY SCHOOLS
9 S GRADE 1	Word proximity, e.g. GRADE(W)1, is not used when searching for grades because the possibility exists for retrieving a substantial number of false drops; GRADE 1
10 S GRADE 2	GRADE 2
11 S GRADE 3	GRADE 3
12 S GRADE 4	GRADE 4
13 S GRADE 5	GRADE 5
14 S GRADE 6	GRADE 6

Search Save (cont.)

FORMULATION	COMMENTS AND DESCRIPTORS/Identifiers
15 S GRADE 7	GRADE 7
16 S GRADE 8	GRADE 8
17 S INT?RMEDIATE(W)GRADE	
18 S INTERMEDIATE(W)GRADES	INTERMEDIATE GRADES
19 S INTERMEDIATE(W)SCHOOL	Intermediate School 201
20 S INTERMEDIATE(W)SCHOOLS	Intermediate Schools
21 S KINDERGART?	KINDERGARTEN, KINDERGARTEN CHILDREN/ Kindergarten, Kindergartner, Kindergartners, 8 multi-term Kindergarter Identifiers
22 S PRIMARY(W)EDUCATION	PRIMARY EDUCATION/Primary Education Project
23 S PRIMARY(W)GRADE	
24 S PRIMARY(W)GRADES	PRIMARY GRADES
25 S PRIMARY(W)PROGRAM	
26 S PRIMARY(W)PROGRAMS	UNGRADED PRIMARY PROGRAMS
27 S PRIMARY(W)SCHOOL	
28 S PRIMARY(W)SCHOOLS	
29 S PRIMARY(W)SYSTEM	NONGRADED PRIMARY SYSTEM
30 C 1-15/OR	COMBINing search statements in ranges is limited so the COMBINE operation is repeated three times in this formulation.
31 C 16-29/OR	
32 C 30-31/OR	

Approximate Processing Time: .182 hrs. ($4.55) in file 1.

SECONDARY EDUCATION

Search Save for: Secondary Education (grades 9-12)

FORMULATION	COMMENTS AND DESCRIPTORS/Identifiers
1 S COLLEGE(W)BOUND	COLLEGE BOUND STUDENTS/College Bound Program, College Bound Programs
2 S COLLEGE(W)PREPARATION	COLLEGE PREPARATION
3 S GRADE 9	GRADE 9
4 S GRADE 10	GRADE 10
5 S GRADE 11	GRADE 11
6 S GRADE 12	GRADE 12
7 S HIGH(W)SCHOOL	12 multi-term HIGH SCHOOL Descriptors, 2 multi-term JUNIOR HIGH SCHOOL Descriptors/ Junior High School, 2 multi-term Junior High School Identifiers
8 S HIGH(W)SCHOOLS	HIGH SCHOOLS, 7 multi-term HIGH SCHOOLS Descriptors, JUNIOR HIGH SCHOOLS
9 S NONCOLLEGE(W)PREPARATORY	NONCOLLEGE PREPARATORY STUDENTS
10 S PRECOLLEGE	Precollege Centers
11 S PRE(W)COLLEGE	Pre College Centers
12 S PREP(W)SCHOOL	
13 S PREPARATORY(W)SCHOOL	
14 S PREPARATORY(W)SCHOOLS	
15 S SECONDARY(W)EDUCATION	SECONDARY EDUCATION, 2 multi-term SECONDARY EDUCATION Descriptors/Secondary Education, Secondary Education Act, Elementary Secondary Education
16 S SECONDARY(W)GRADE	
17 S SECONDARY(W)GRADES	SECONDARY GRADES

Search Save: (cont.)

FORMULATION	COMMENTS AND DESCRIPTORS/Identifiers
18 S SECONDARY(W)SCHOOL	5 multi-term SECONDARY SCHOOL Descriptors/ Secondary School, 6 multi-term Secondary School Identifiers
19 S SECONDARY(W)SCHOOLS	SECONDARY SCHOOLS, BIRACIAL SECONDARY SCHOOLS
20 C 1-19/OR	

Approximate Processing Time: .170 hrs. ($4.25) in file 1.

DIALOG® Information Retrieval Service

ONTAP® ERIC FILE 201

SAMPLE SEARCH QUESTIONS

Question Number	Topic
SIMPLE SEARCHES	
S02	FID (International Federation for Documentation) and Library or information networking
S03	4-H Clubs, their members, and their activities
S04	Revision of the Anglo-American Cataloging Rules
S05	Navaho language textbooks or grammars (material written in Navaho language, or useful for teaching the Navaho language or about Navaho linguistics)
S06	Education in Sri Lanka (including library activities)
S07	Publications about literacy in developing countries, that are published by the International Institute for Adult Literacy Methods (or any publications by or about this Institute)
S08	16 Personality Factor Test
S09	Parapsychology
MEDIUM-COMPLEXITY SEARCHES	
M01	Direct charging to users for reference and current awareness service of libraries and other information service agencies (philosophy, policy, practice, fees, charges; for any type of library, for any type of reference service; not interested in free services)
M02	Federal aid to day care centers or services for children of working or student parents (including history or philosophy, needs, arguments for and against, reported experiences, evaluations, funding, parent involvement and attitude; restrict to U.S. centers or services)
M03	Jean Piaget's theories and the thought processes or language development of children, but limited to those publications that are available from ERIC/EDRS
M04	Libraries and librarians in the Middle East, particularly in Arabic speaking countries (exclude African countries except for Egypt, include all types of libraries or information centers)
M05	Library service to the physically handicapped (not mentally or language-handicapped)
M06	Effects of TV violence on children
M07	Drug abuse, including alcohol, among students of elementary or secondary schools, grades K-12 (including school education programs on this topic, and sociological studies)

M08	Use of school busing to achieve racial integration
M09	Recreational use of forest lands
M10	Title 9 of the 1972 Federal Education Act Amendments, and women's sports
M11	White flight to the suburbs

DIFFICULT SEARCHES

D01	Training for supervision and management in libraries and information centers (including the need for such training descriptions of training programs or materials; training of library school students and professional working librarians; formal academic training as well as on the job training)
D02	Audiovisual aids for orientation or instruction of library users
D03	Evaluation of primary school (grades K-3) English reading programs or reading materials and techniques (but not the evaluation of specific reading tests or instructors, and not the student test scores when they are not being used as part of an evaluation of the reading program and not just the criteria or standards for evaulation). Limit the output to publications available from ERIC/EDRS
D04	Formal science education programs in universities and secondary schools (grades 9-12) in the Soviet Union (all sciences, including mathematics and engineering)
D05	Vocational education of the American Indian (history, data, and programs to provide this education; but not training materials to be used in these education programs)
D06	Evaluation of bilingual elementary (grades K-8) and secondary (grades 9-12) school programs and techniques, specifically those that involve both Spanish and English languages
D07	The financial costs to institutions of higher education to implement legislated affirmative action programs
D08	Evaluation of indexing or cataloging (work, methods, products, languages). This is to include the representation and storage of the index information. Indexing is defined broadly to include all forms of text searching, i.e., indexing by text words. Evaluation is to be interpreted in a broad sense such as time and cost, error rates, recall/precision, and criteria for evaluation. Indexing is to include generic processes such as citation indexing, title word indexing, or automatic indexing. The papers must contain some evaluative aspect; a description of an index or an index or indexing system by itself is not enough.
D09	Collective bargaining in libraries of institutions of higher education

ERIC/IR
Special Project Report

ONLINE SEARCHING OF ERIC

PART IV

THE REDESIGN OF THE ERIC DATA BASE FOR ONLINE SEARCHING

ERIC Clearinghouse on Information Resources
Syracuse University
1979

PART IV

THE REDESIGN OF THE ERIC DATA BASE FOR ONLINE SEARCHING

by Pauline Atherton and Karen Markey

Abstract

The format of items in the two ERIC printed services (RIE and CIJE) are compared with the online version(s) of these records which exist on LRS, SDC, and BRS. These differences and those in the vocabulary tools (the basic indexes and the thesaurus) are shown to point up possible causes for confusion to the users. Several suggestions for improvements are made--some to do with standardization, others to do with new features for searching the ERIC vocabulary online, for printing hits, and for producing searches more cost-effectively. A review of previous retrieval effectiveness studies ends the report.

Contents

	Page
INTRODUCTION	1
I. WHAT IS ERIC ONLINE?	3
A. Online Printed ERIC Records	7
B. The Order of ERIC Postings Online	11
C. The ERIC Vocabulary--In Print and Online	14
D. Recommendations for the Online ERIC Record	19
E. Specific Recommendations for ERIC Online Basic Index and Online Thesaurus	19
II. HOW IS ERIC SEARCHED ONLINE?	22
A. Background	22
B. Recommendations for Searching ERIC Online	30
III. WHAT ABOUT THE ERIC ONLINE USER?	31
A. Background	31
B. General Recommendations	34
C. Specific Recommendations from Online Searchers	37
D. Comments on Searching Features	39

Contents (Con't.)

Page

E. Comments With Suggestions for Improvement of Vocabulary Browsing Features 44

1. Need to improve the ability to select from the online dictionary and thesaurus. 44
2. Need for an online rotated descriptor display 47
3. Need for an online rotated identifier display 47
4. Need for a free text wrap-around display. 48
5. Need for Lead-In from free text to controlled vocabulary. . . . 51
6. Highlighting search term lists. 52
7. Changes to existing print functions 54
8. Expanded search save capabilities 56
9. String searching. 56
10. Stop word lists 57
11. Search site profiles. 57
12. Housekeeping functions. 58
13. Retrieval system improvement. 58
14. For the data base 59

IV. FREE TEXT OR CONTROLLED VOCABULARY SEARCHING--A REVIEW OF PREVIOUS WORK WITH A POSSIBLE BEARING ON REDESIGN. 62

V. RECOMMENDATIONS FOR FUTURE STUDY 70

Figures, Tables and Illustrations

Figure 1. ERIC Unit Record BRS/LRS/SDC 4
Figure 2. The Inverted File for ERIC Online LRS, SDC and BRS. 5
Figure 3. ERIC Unit Record BRS/LRS/SDC Print Qualifiers. 6
Figure 4. Psychological Abstracts' Record Reformatted to Resemble ERIC format for WEECN 35

Bates
Table 1. Summary of Information Search Tactics and Definitions 26

Bates
Table 2. Suggested Tactics in Response to Requests 28

Table 3. Analysis of Existing Features in Retrieval Systems. 38
Table 4. Comparison of Retrieval System Commands for Searching Printing and Housekeeping Tasks 41

Illustration A. CIJE Record In Print and on BRS, SDC, LRS 8
Illustration B. RIE Record In Print and on BRS, SDC, LRS. 9
Illustration C. In Print Indexes in RIE and CIJE. 12
Illustration D. RIE Author Index and LRS Author Printout. 13
Illustration E. ERIC Thesaurus--In Print and Online on LRS. 15
Illustration F1. In Print Rotated Descriptor Display and Online Browsing in Basic Index on LRS, SDC and BRS. 16
Illustration F2. Browsing in Basic Index of ERIC Online on LRS, SDC and BRS. 17

THE REDESIGN OF THE ERIC DATA BASE FOR ONLINE SEARCHING

Introduction

Bibliographic reference retrieval systems may not be at a crossroad, but the insight we gained from this study has caused us to pause and wonder where we can go from here. ERIC, as a database, was one of the first to be put online. Its history spans the complete historic period of the "online revolution." As a leader, a pioneer, it has and probably will continue to be an exemplar for other data base producers getting into the online business. We hope that our examination of the uses of ERIC online and our recommendations for improvement of the data base to accommodate online users will be read by all data base producers.

To pull together many of the suggestions we ourselves have made or heard from "the field," we have written the following part of the report. It starts with illustrations of ERIC in print and online to show differences (some tolerable, others questionable) which may confuse ERIC users. We move from these illustrations into the world of online searching--retrieval system features which cause problems, could be improved, and from there to the world of the future online systems, suggesting new features.

All of this part is offered in the hope that someone "out there" can effect improvements--either of the data bases and its indexes or of the retrieval systems. As diffuse as the control is, the world of online searching needs some standardization, some systematic coordination, and some simplification if we can ever hope to make online bibliographic searching a task any user can perform him/herself.

People like Charlie Meadow* and Linda C. Smith**have made suggestions which we should heed as advances in the art of online searching. Redesign of data bases is one place to start, but it can not be done in the vacuum of one information system any longer. National and international networks for information services are the order of the day for the 1980's.

---

*Meadow, C.T., Hewett, T.T., Rafsnider, D.J., Toliver, D.E., Epstein, B., Edelmann, J.V., and Maher, A. Individualized Instruction for Data Access (IIDA). Final Design Report. Philadelphia: Drexel University, 1977.

**Smith, L.C. Artificial Intelligence in Information Retrieval Systems. (Doctoral dissertation, Syracuse University, 1979)

I. What is ERIC Online?

In print, ERIC provides two literature-searching tools, the two monthly abstracting journals, CIJE and RIE. An auxiliary tool is the 7th edition of the Thesaurus of ERIC Descriptors (or one of the semi-annual "working copies" of this thesaurus).

In machine-readable form, at the ERIC Facility in Bethesda, ERIC is a Master Tape of two files, CIJE and RIE, an inverted file of descriptors and identifiers (major and minor), and a file of the Thesaurus of ERIC Descriptors with their display of related terms.

ERIC Online, on any of several commercial retrieval system vendors (Lockheed's is DIALOG, System Development Corporation's ORBIT IV, BRS), or other academic or private retrieval systems, is a data base -- a set of files:

1. the file of ERIC records -- the Master Tape from the ERIC Facility, reformatted with data element codes changed to match the "standard" for each vendor's or system's set of files (see Figure 1).

2. the inverted file -- sometimes referred to as the dictionary or basic index -- this contains directly searchable single word terms from several fields of the ERIC record (see Figure 2). This points the computer back to the records containing these words. Often the controlled vocabulary/ indexing record from the descriptor and identifier fields of the ERIC records are included in this Basic Index, as they appear, complete with a * for major descriptors. This file is directly searchable and browsable in all online systems with no code designators required.

3. the lists of terms in fields searchable only by code designators, e.g., AU = Jones, YR = 78, etc., (see Figure 1).

4. the file of the ERIC Thesaurus. Now only commercially available on DIALOG.

Each online commercial vendor determines which fields are printable (see Figure 3). They also decide how these fields will be arranged, labeled or explained in printouts.

Figure 1. ERIC UNIT RECORD - BRS/LRS/SDC Search Qualifiers

ERIC Field Names*	Master Tape ERIC Codes (RIE/CIJE)	BRS-Search Qualifier	LRS-Search Qualifier	SDC-Search Qualifier
Accession Number	0010	.AN.	---	/AN
Clearinghouse Accession Number	CH 0011	---	---	/CHAN
Clearinghouse Code		.CH.	CH=	/CC
Title*	001A	.TI.	/TI	/TI
Author*	AUTH 001B	.AU.	AU=	/AU
Corporate Source*	INST 001C	.IN.	/CS	/OS
Sponsoring Agency	SPON 0020	.SN.	/SA	/SPO
Descriptor*	DESC 0023	.MJ,MN.	/DE or /DF or /DE* or /DF*	/IT
Identifier*	IDEN 0024	.ID.	/ID or /IF or /ID* or /IF*	/IT
Abstract*	ABST 002C	.AB.	/AB	/BI
Descriptive Note	NOTE 0026	---	/NT	---
Publication Year*	PDAT 0017	.YR.	YR=	FROM__ TO__
Pagination*	PAGE			
Journal Name*	JNL 0032	---	JO=	/JC
Legislative Authority Code		.PA.	AC=	/LPA
Contract/Grant Number	GR 002F CONT 002E	---	CN=	/NU
Bureau/Project Number	BN 0030	---	PN=	/NU
Report Number*	REPNO 002D	---	RN=	/NU
Corporate Source Code	0080	.IN.	SC=	---
Sponsoring Agency Code	0084	.SN.	SC=	---
Document Type*	PUBTYPE 0013	.PT.	DT=	/DT
Group Code	0014(?)	---	GC=	---
EDRS Price	0025	---	---	---
Availability Level	AVAIL 0031	---	---	---
Update	0001	---	UD=	/UP
Issue	002B	---	IS=	/IS

*Several of the fields in the ERIC record are listed in the "Recommended Interchange Specifications for Computer-Readable Bibliographic Data Bases" (April 1978 publication of ASIDIC-EUSIDIC-ICSU-AB-NFAIS). Those found in their list of essential data elements are marked here with an asterisk (*). Their field names vary from ERIC, e.g., Journal is Serial Title, Title is Title of contribution, Publication Date is Date of issue or imprint, etc.

As of January, 1979, ERIC added three new fields (GOV, LANG, and GEO) to the RIE & CIJE Master Tape. These are not in this table because the respective systems have not yet indicated whether or not they will be searchable directly or by prefix codes.

Figure 2. The Inverted File (i.e., Basic Index) for ERIC Online LRS, SDC and BRS

A. Directly Searchable Fields:	Field Name			
Basic Index Contains Single Words from:	Author			AU
	Institution			IN
	Title	TI	TI	TI
	Spon. Agency	SA[b]		SN
	Corp. Source	CS[b]		
	Descriptor	DE	IT	MJ, MN
	Identifier	ID	ST	ID
	Note	NT		
	Abstract	AB	AB	AB

a) In descriptor and identifier fields of BRS record (MJ, MN) only single word descriptors are listed in Basic Index. Single words within multi-word descriptors are not accessed, but plans are underway to change this.

b) LRS plans to delete these fields from Basic Index at time of next update.

Figure 3. ERIC Unit Record BRS/LRS/SDC Print Qualifiers

Data Element Name	BRS-Print Qualifier	Included LRS-Print format	SDC-Print Qualifier
Accession Number	AN	1-6	AN
Clearinghouse Accession Number	----	2,5	CHAN
Clearinghouse Code	CH	2,5	----
Title	TI	2,6	TI
Author	AU	2,3,5	AU
Corporate Source	IN	2,3,5	OS
Sponsoring Agency	SN	2,5	SPO
Descriptor	MJ,MN	2,5	IT
Identifier	ID	2,5	ST
Abstract	AB	4,5	AB
Descriptive Note	NT	2,3,5	NO
Publication Year	YR	2,3,5	SO
Journal Name	SO	2,3,5	SO
Legislative Authority Code	PA	2,5	LPA
Contract/Grant No.	NO	2,5	NU
Bureau/Project No.	NO	2,5	NU
Report Number	NO	2,5	NU
Corporate Source Code	IN	2,5	----
Sponsoring Agency Code	SN	2,5	----
Document Type	PT	2,5	DT
Group Code	----	----	----
EDRS Price	PR	2,3,5	AV
Availability	AV	2,3,5	AV
Update	----	----	UP
Issue	IS	----	IS

A. Online Printed ERIC Records

From the following illustrations, one can compare the order of the fields in the ERIC record as it appears in print, in CIJE and RIE, and as it appears "in print" online in the BRS, LRS, and SDC systems. Use Figure 3 as a reference.

CIJE (See Illustration A)

BRS and SDC give the data element codes opposite each field; LRS does not. Many data elements are rearranged in one or the other systems, if CIJE's order is used as the standard.

1. BRS prints author ahead of title and places CH code after source or journal title. Availability is placed before year, not at end of annotation. Descriptors are rearranged to list MJ separately from MN and ID, with no commas separating each, and multiple term descriptors "bundled" by hyphens. Annotator data are deleted from BRS version of CIJE record, as is "v" "n" and "pp" from journal citation.

2. SDC gives all information except availability information. The record is slightly reformatted, with "v" "n" and "pp" deleted from journal citation. SDC includes CIJE issue date. (Note that descriptors are IT and identifiers are coded ST.)

3. LRS gives the most exact rendition of all CIJE information, with most readable format: indentation for bibliographic information, followed by annotation, followed by "explained" fields of descriptors and identifiers.

RIE (See Illustration B)

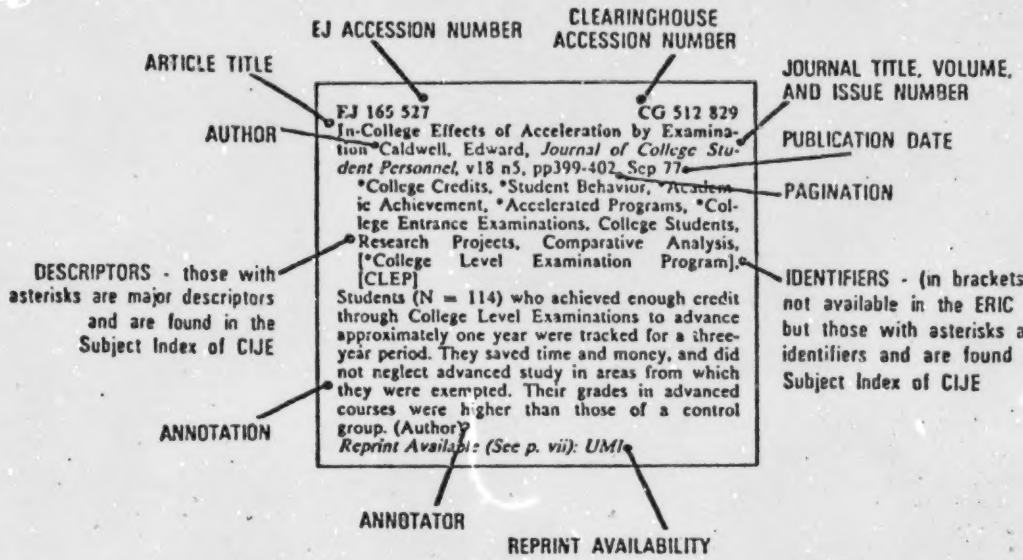
The RIE record is more complicated. The fields in common with CIJE will not be reviewed again in this comparison across BRS, SDC, and LRS versions of ERIC Online records.

Attention is focused on the treatment of additional fields:

1. BRS includes the codes for agencies in their IN and SN fields. "Pub Date" becomes "SO." YR field is added. PR is placed above AV and NT--a very different order from RIE.

2. SDC places title before author. They mysteriously added "IPP" to SO field. AV is before NO field. They have added the field DT and explained the code.

Illustration A. CIJE Record in Print and on BRS, SDC, LRS


EJ ACCESSION NUMBER
CLEARINGHOUSE ACCESSION NUMBER
ARTICLE TITLE
JOURNAL TITLE, VOLUME, AND ISSUE NUMBER
AUTHOR
PUBLICATION DATE
PAGINATION
DESCRIPTORS - those with asterisks are major descriptors and are found in the Subject Index of CIJE
IDENTIFIERS - (in brackets) are not available in the ERIC Thesaurus, but those with asterisks are major identifiers and are found in the Subject Index of CIJE
ANNOTATION
ANNOTATOR
REPRINT AVAILABILITY
EJ 165 527 CG 512 829
In-College Effects of Acceleration by Examination. Caldwell, Edward, Journal of College Student Personnel, v18 n5, pp399-402, Sep 77.
*College Credits, *Student Behavior, *Academic Achievement, *Accelerated Programs, *College Entrance Examinations, College Students, Research Projects, Comparative Analysis, [*College Level Examination Program], [CLEP]
Students (N = 114) who achieved enough credit through College Level Examinations to advance approximately one year were tracked for a three-year period. They saved time and money, and did not neglect advanced study in areas from which they were exempted. Their grades in advanced courses were higher than those of a control group. (Author)
Reprint Available: (See p. vii): UMI

BRS

```
BRS - SEARCH MODE - ENTER QUERY
   2 :   EJ165527.AN
   RESULT       1

   3 :   ..PRINT 2 ALL/DOC=1
     1
AN  EJ165527
AU  CALDWELL, EDWARD.
TI  IN-COLLEGE EFFECTS OF ACCELERATION BY EXAMINATION
SO  JOURNAL OF COLLEGE STUDENT PERSONNEL.
    18. 5. 399-402 SEP 77.
IS  CIJE1978.
CH  CG512829.
AV  REPRINT AVAILABLE (SEE P. VII): UMI.
YR  77
MJ  COLLEGE-CREDITS STUDENT-BEHAVIOR ACADEMIC-
    ACHIEVEMENT ACCELERATED-PROGRAMS COLLEGE-ENTRANCE-
    EXAMINATIONS.
MN  COLLEGE-STUDENTS RESEARCH-PROJECTS COMPARATIVE-
    ANALYSIS.
ID  COLLEGE LEVEL EXAMINATION PROGRAM: CLEP.
AB  STUDENTS (N = 114) WHO ACHIEVED ENOUGH CREDIT
    THROUGH COLLEGE LEVEL EXAMINATIONS TO ADVANCE
    APPROXIMATELY ONE YEAR WERE TRACKED FOR A
    THREE-YEAR PERIOD. THEY SAVED TIME AND MONEY,
    AND DID NOT NEGLECT ADVANCED STUDY IN AREAS
    FROM WHICH THEY WERE EXEMPTED. THEIR GRADES IN
    ADVANCED COURSES WERE HIGHER THAN THOSE OF A
    CONTROL GROUP.
```

SDC

```
SS 2 /C?
USER:
"FIND EJ165527/AN""PRT FULL"

PROG:
SS 2PSTG (1)

-1-
AN  - EJ165527
CHAN- CG512829
TI  - IN-COLLEGE EFFECTS OF ACCELERATION BY EXAMINATION
AU  - CALDWELL, EDWARD
SO  - JOURNAL OF COLLEGE STUDENT PERSONNEL: 18: 5:
      399-402 (SEP 1977)
IS  - CIJE78
IT  - *ACCELERATED PROGRAMS; *ACADEMIC ACHIEVEMENT;
      *COLLEGE ENTRANCE EXAMINATION; *COLLEGE CREDITS;
      COMPARATIVE ANALYSIS; COLLEGE STUDENTS; RESEARCH
      PROJECTS; *STUDENT BEHAVIOR
ST  - *COLLEGE LEVEL EXAMINATION PROGRAM: CLEP
AN  - STUDENTS (N = 114) WHO ACHIEVED ENOUGH CREDIT
      THROUGH COLLEGE LEVEL EXAMINATIONS TO ADVANCE
      APPROXIMATELY ONE YEAR WERE TRACKED FOR A THREE-
      YEAR PERIOD. THEY SAVED TIME AND MONEY, AND
      DID NOT NEGLECT ADVANCED STUDY IN AREAS FROM
      WHICH THEY WERE EXEMPTED. THEIR GRADES IN
      ADVANCED COURSES WERE HIGHER THAN THOSE OF A
      CONTROL GROUP.  (AUTHOR)
```

LRS

```
? TEJ165527/5
EJ165527/5
EJ165527 CG12829
  IN-COLLEGE EFFECTS OF ACCELERATION BY EXAMINATION
  CALDWELL, EDWARD
  JOURNAL OF COLLEGE STUDENT PERSONNEL, 18. 5. 399-402
  SEP 77
  REPRINT AVAILABLE (SEE P. VII): UMI
STUDENTS (N = 114) WHO ACHIEVED ENOUGH CREDIT THROUGH
COLLEGE LEVEL EXAMINATIONS TO ADVANCE APPROXIMATELY
ONE YEAR WERE TRACKED FOR A THREE-YEAR PERIOD. THEY
SAVED TIME AND MONEY, AND DID NOT NEGLECT ADVANCED
STUDY IN AREAS FROM WHICH THEY WERE EXEMPTED. THEIR
GRADES IN ADVANCED COURSES WERE HIGHER THAN THOSE OF
A CONTROL GROUP.  (AUTHOR)
  DESCRITTORS: *COLLEGE CREDITS/ *STUDENT BEHAVIOR/
*ACADEMIC ACHIEVEMENT/ *ACCELERATED PROGRAMS/ *COLLEGE
ENTRANCE EXAMINATIONS/ COLLEGE STUDENTS/ RESEARCH
PROJECTS/ COMPARATIVE ANALYSIS
  IDENTIFIERS: *COLLEGE LEVEL EXAMINATION PROGRAM/ CLEP
```

Illustration B. RIE Record in Print and on BRS, SDC, LRS

BRS

```
BRS - SEARCH MODE - ENTER QUERY
1_:   ED148874.AN.
RESULT
2_: ..PRINT 1/ALL/DOC=1
1
AN  ED148874
AU  JACKSON, REX.
IN  COLLEGE ENTRANCE EXAMINATION BOARD, NEW YORK,
    N.Y.  (QPX15900).
TI  A SUMMARY OF SAT SCORE STATISTICS FOR COLLEGE
    BOARD CANDIDATES.
SO  OCT 76.
SN  EDUCATIONAL TESTING SERVICE, PRINCETON, N.J.
    (QAT24225).
IS  RIEMAY78.
CH  TM006764.
PR  EDRS PRICE MF-$0.83 PLUS POSTAGE.  HC NOT
    AVAILABLE FROM EDRS.
AV  COLLEGE BOARD PUBLICATION ORDERS, BOX 2815,
    PRINCETON, NEW JERSEY  08541 (ITEM NO.
    2517183, $2.00).
NT  32P.  FOR RELATED DOCUMENTS, SEE TM 006 750-
    776 and TM 006 805; FOR COMPLETE REPORT, SEE
    TM 006 516.
YR  77.
MJ  COLLEGE-ENTRANCE-EXAMINATIONS EDUCATIONAL-
    TRENDS NORMS SCORES STUDENT-CHARACTERISTICS.
MN  ACADEMIC-ACHIEVEMENT ACHIEVEMENT-RATING
    COLLEGE-BOUND-STUDENTS COLLEGE-MAJORS
    EDUCATIONAL-BACKGROUND EDUCATIONAL-PROBLEMS
    GEOGRAPHIC-REGIONS HIGH-SCHOOL-STUDENTS
    PERFORMANCE-FACTORS QUESTIONNAIRES SECONDARY-
    EDUCATION SEX-DIFFERENCES TABLES-DATA TESTING-
    PROBLEMS TEST-RESULTS TREND-ANALYSIS.
ID  CLASS RANK; SCHOLASTIC APTITUDE TEST; TEST SCORE
    DECLINE.
AB  SUMMARIES OF SCHOLASTIC APTITUDE TEST (SAT) SCORE
    STATISTICS (MEANS, STANDARD DEVIATIONS, AND THE
    NUMBER OF STUDENTS RAKING THE TEST) ARE PRESENTED
    FOR STUDENTS GROUPED ACCORDING TO TESTING YEARS
    (FOR 1956-57 TO 1975-76 AND ACCORDING TO...
```

SDC

```
SS 1 /C7
USER:
"FIND ED148874/AN""PRT FULL"

PROG:
SS 1 PSTG (1)

-1-
AN   - ED148874
CHAN - TM006764
TI   - A SUMMARY OF SAT SCORE STATISTICS FOR COLLEGE
       BOARD CANDIDATES
AU   - JACKSON, REX
OS   - COLLEGE ENTRANCE EXAMINATION BOARD, NEW YORK,
       N.Y.
SPO  - EDUCATIONAL TESTING SERVICE, PRINCETON, N.J.
SO   - OCT 1976; 1PP
IS   - RIE78MAY
AV   - EDRS PRICE MR-$0.83 PLUS POSTAGE.  HC NOT
       AVAILABLE FROM EDRS.  932 MF).  COLLEGE BOARD
       PUBLICATION ORDERS, BOX 2815, PRINCETON, NEW
       JERSEY  08541 (ITEM NO. 2517183, $2.00)
NO   - FOR RELATED DOCUMENTS, SEE TM 006 750-776 AND
       TM 006 805; FOR COMPLETE REPORT, SEE TM 006 516
DT   - R (RESEARCH & TECHNICAL REPORTS)
IT   - ACADEMIC ACHIEVEMENT; ACHIEVEMENT RATING; COLLEGE
       BOUND STUDENTS; *COLLEGE ENTRANCE EXAMINATIONS;
       COLLEGE MAJORS; EDUCATIONAL BACKGROUND; EDUCATIONAL
       PROBLEMS; *EDUCATIONAL TRENDS; GEOGRAPHIC REGIONS;
       HIGH SCHOOL STUDENTS; *NORMS; PERFORMANCE FACTORS;
       QUESTIONNAIRES; *SCORES; SECONDARY EDUCATION;
       SEX DIFFERENCES; *STUDENT CHARACTERISTICS;
       TABLES (DATA); TESTING PROBLEMS; TEST RESULTS;
       TREND ANALYSIS
ST   - CLASS RANK; *SCHOLASTIC APTITUDE TEST; *TEST
       SCORE DECLINE
AB   - SUMMARIES OF SCHOLASTIC APTITUDE TEST (SAT)
       SCORE STATISTICS (MEANS, STANDARD DEVIATIONS,
       AND THE NUMBER OF STUDENTS TAKING THE TEST)
       ARE PRESENTED FOR STUDENTS GROUPED ACCORDING TO
       TESTING YEARS (FOR 1956-57 TO 1975-76 AND
       ACCORDING TO HIGH SCHOOL CLASSES...
```

RIE

ED 148 874 **TM 006 764**

Jackson, Rex

A Summary of SAT Score Statistics for College Board Candidates.

College Entrance Examination Board, New York, N.Y.

Spons Agency—Educational Testing Servi Princeton, N.J.

Pub Date Oct 76

Note—32p.; For related documents, see TM 006 750-776 and TM 006 805; For complete report, see TM 006 516

Available from—College Board Publication Orders, Box 2815, Princeton, New Jersey 08541 (Item No. 2517183, $2.00)

EDRS Price MF-$0.83 Plus Postage. HC Not Available from EDRS.

Descriptors—Academic Achievement, Achievement Rating, College Bound Students, *College Entrance Examinations, College Majors, Educational Background, Educational Problems, *Educational Trends, Geographic Regions, High School Students, *Norms, Performance Factors, Questionnaires, *Scores, Secondary Education, Sex Differences, *Student Characteristics, Tables (Data), Testing Problems, Test Results, Trend Analysis

Identifiers—Class Rank, *Scholastic Aptitude Test, *Test Score Decline

Summaries of Scholastic Aptitude Test (SAT) score statistics (means, standard deviations, and the number of students taking the test) are presented for students grouped acco[illegible] ing years (fo[illegible] from 1971-[illegible] are similar in terms of ethnic background, rank in class, and educational goals. There is, however, a trend away from liberal arts to other fields--primarily career-oriented fields such as business, education, and health. (Author/EVH)

LRS

```
ED148874/5
ED148874 TM006764
  A SUMMARY OF SAT SCORE STATISTICS FOR COLLEGE BOARD
CANDIDATES.
  JACKSON, REX
  COLLEGE ENTRANCE EXAMINATION BOARD, NEW YORK, N.Y.
  OCT 76  32P.; FOR RELATED DOCUMENTS, SEE TM 006
750-776 AND TM 006 805; FOR COMPLETE REPORT, SEE
TM 006 516
  SPONSORING AGENCY:  EDUCATIONAL TESTING SERVICE,
PRINCETON, N.J.
  AVAILABLE FROM:  COLLEGE BOARD PUBLICATION ORDERS,
BOX 2815, PRINCETON, NEW JERSEY  08541 (ITEM NO.
2517183, $2.00)
  EDRS PRICE MF-$).83 PLUS POSTAGE.  HC NOT AVAILABLE
FROM EDRS.
  SUMMARIES OF SCHOLASTIC APTITUDE TEST (SAT) SCORE
STATISTICS (MEANS, STANDARD DEVIATIONS, AND THE NUMBER
OF STUDENTS TAKING THE TEST) ARE PRESENTED FOR STUDENTS
GROUPED ACCORDING TO TESTING YEARS (FOR 1956-57 TO
1975-76) AND ACCORDING TO HIGH SCHOOL CLASSES...
  DESCRIPTORS:  ACADEMIC ACHIEVEMENT/ ACHIEVEMENT RATING/
COLLEGE BOUND STUDENTS/ *COLLEGE ENTRANCE EXAMINATIONS/
COLLEGE MAJORS/ EDUCATIONAL BACKGROUND/ EDUCATIONAL
PROBLEMS/ *EDUCATIONAL TRENDS/ GEOGRAPHIC REGIONS/
HIGH SCHOOL STUDENTS/ *NORMS/ PERFORMANCE FACTORS/
QUESTIONNAIRES/ * SCORES/ SECONDARY EDUCATION/ SEX
DIFFERENCES/ *STUDENT CHARACTERISTICS/ TABLES (DATA)/
TESTING PROBLEMS/ TEST RESULTS/ TREND ANALYSIS
  IDENTIFIERS:  CLASS RANK/ *SCHOLASTIC APTITUDE TEST/
*TEST SCORE DECLINE
```

3. LRS explains several fields (e.g., sponsoring agency, available from). They do not include DT field in this printed version, but it is a code-searchable field.

As the years of online searching have progressed, the difference between BRS, SDC, and LRS versions of the ERIC Master Tape and related files have become more apparent to ERIC searchers. Searchers have made requests to online vendors which have changed the structure of some of these fields. For example, DIALOG will revise their Basic Index for ERIC, deleting corporate source and sponsoring agency data elements because it has caused confusion for online searchers. BRS plans to unbundle descriptors in their Basic Index to allow searchers to access records by any of the words in a multi-word descriptor.

Each commercial online vendor probably updates their ERIC databases (all files) monthly.

Occasionally they change or add to the format of the record printed. As we can see from the SDC record, they have recently added the DT field. All vendors will need to make decisions about adding to their formats the new GOV, GEO, and LAN fields of the ERIC record introduced in January, 1979.

Most will need to be prepared to upgrade their version of ERIC when the Identifier Clean-Up project is finished a d the total ERIC Master Tape is upgraded after the Vocabulary Improvement Project is completed.

These differences are of some concern to us as they were to Oldroyd and Citroen* (1977). These researchers noted that searchers seemed to pay little attention to differences in data base characteristics and the adaptation necessary because of these differences.

---

*B.K. Oldroyd and C.L. Citroen, "Study of Strategies in Online Searching," Online Review, 1 (1977), 295-310.

B. The Order of ERIC Postings Online

Searching a single descriptor or a single author online is similar to creating a subject index or author index similar to CIJE and RIE Author and Subject Indexes. See Illustration C for an example of the inprint ERIC indexes.

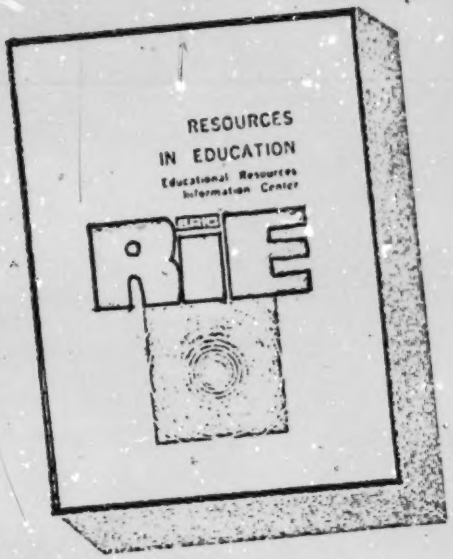
From Illustration D, one can see the differences in such indexes, in print and online.

Online, if one uses the LRS/DIALOG system, and limits the search to the ED portion of the file, the printout is in reverse order of ED number (i.e., most recent first).

Note: An added benefit online is that a cumulative index printout is possible. (Five more citations by Rex Jackson are in the ERIC/LRS file but are not shown in the illustration.)

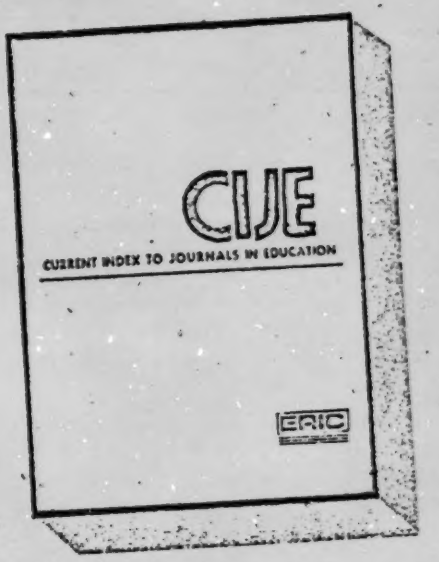
We might ask, "Which is better?" "Why should they be different?" Some questions are worth asking but are not easy to answer. We only note differences here to check on their need to exist as possible confusion points for the ERIC searcher, be he at the printed index or online.

Illustration C. In Print Indexes in RIE and CIJE



INDEXES
SUBJECT
AUTHOR
INSTITUTION

INDEXES
SUBJECT
AUTHOR
JOURNAL
CONTENTS



JOURNAL CONTENTS:

Journal of College Student Personnel *Sep 77*

Characteristics of Student Judicial Board Members: Their Relationship to Disciplinary Decision Making EJ 165 517
Crime on the Campus EJ 165 518
Group Relations Training and Its Relevance for the Counselor EJ 165 519
Differences in the Utilization of Professional and Paraprofessional Counseling Services EJ 165 520
Factors Associated with Postcollege Changes in Personality Characteristics EJ 165 521
Psychological Discomfort and Its Causal Attribution in Relation to Student Services Program Development EJ 165 522
Student Services in Five Alberta Colleges: A Measure of Quality EJ 165 523
A Comparison of the Adjustment Problems of Three International Student Groups EJ 165 524
Restructuring Residence Hall Programming: Residence Hall Educators with a Curriculum EJ 165 525
Group Assertive Training for Resident Assistants EJ 165 526
In-College Effects of Acceleration by Examination EJ 165 527
Developmental Education: The Recognition of Individual Differences EJ 165 528
Counseling Needs of Disadvantaged Veterans EJ 165 529
The Challenge of Student Development and Steady State Issues in Higher Education EJ 165 530
On Modeling Reality EJ 165 531
The Student Cultures: The Challenge of Privatism EJ 165 532
Some Perspectives on Carts and Horses in Student Development EJ 165 533

SUBJECT:

College Entrance Examinations
In-College Effects of Acceleration by Examination, *Journal of College Student Personnel* v18 n5, pp399-402, Sep 77
EJ 165 527

AUTHOR:

Caldwell, Edward
In-College Effects of Acceleration by Examination EJ 165 527

Institution Index

Coalition of Indian Controlled School Boards, Inc., Denver, Colo.
Proceedings: Conference on "Opportunities in Vocational Education for American Indians" (Denver, Colorado, June 28-30, 1977).
ED 147 536

College Entrance Examination Board, New York, N.Y.
An Analysis of Textbooks in Relation to Declining SAT Scores.
ED 148 865
Annotated Bibliography.
ED 148 879
Available Evidence on Public Attitudes toward Education. With a supplement of excerpts from Eighth Annual Survey of the Public's Attitudes toward the Public Schools.
ED 148 878
Changes in the Verbal Abilities of High School Seniors, College Entrants, and SAT Candidates between 1960 and 1972.
ED 148 861

Illustration D.

RIE Author Index

and LRS Author Printout

Author Index

Jackson, Gregory A.
Description of Merged Data Base: Appendix F. The Development of Institutions of Higher Education. Theory and Assessment of Impact of Four Possible Areas of Federal Intervention.
ED 148 240
Financial Aid and Student Enrollment: Appendix E. The Development of Institutions of Higher Education: Theory and Assessment of Impact of Four Possible Areas of Federal Intervention.
ED 148 241

Jackson, Rex
Comparison of SAT Score Trends in Selected Schools Judged to Have Traditional or Experimental Orientations.
ED 148 870
Correlations of SAT Scores with High School Record.
ED 148 871
An Examination of Declining Numbers of High-Scoring SAT Candidates.
ED 148 872
Mean Scores for PSAT and PSAT/NMSQT Junior Candidates.
ED 148 873
A Summary of SAT Score Statistics for College Board Candidates.
ED 148 874
Trends in Mean SAT Scores for Students from the High School Classes of 1971-72 and 1975-76 Belonging to Selected School and College Groups.
ED 148 875

Jacob, Jasmine Fernando
The Effects of Individualizing Instruction by the Use of Multi-Sensory Multi-Media Learning Centers on Reading and Arithmetic Achievement of Inner-City Children.
ED 147 765//

LRS

```
S AU=JACKSON, REX?
          1    12 AU=JACKSON, REX ?
? L1/ED
          2    11 1/ED
? T /6/1-11  (only six are shown)
2/6/1
ED148875
  TRENDS IN MEAN SAT SCORES FOR STUDENTS FROM THE HIGH SCHOOL CLASSES
OF 1971-72 AND 1975-76 BELONGING TO SELECTED SCHOOL AND COLLEGE GROUPS.

2/6/2
ED148874
  A SUMMARY OF SAT SCORE STATISTICS FOR COLLEGE BOARD CANDIDATES.

2/6/3
ED148873
  MEAN SCORES FOR PSAT AND PSAT/NMSQT JUNIOR CANDIDATES.

2/6/4
ED148872
  AN EXAMINATION OF DECLINING NUMBERS OF HIGH-SCORING SAT CANDIDATES.

2/6/5
ED148871
  CORRELATIONS OF SAT SCORES WITH HIGH SCHOOL RECORD.

2/6/6
ED148870
  COMPARISON OF SAT SCORE TRENDS IN SELECTED SCHOOLS JUDGED TO HAVE
TRADITIONAL OR EXPERIMENTAL ORIENTATIONS.
```

C. The ERIC Vocabulary--In Print and Online

In print, The Thesaurus of ERIC Descriptors lists lead-in terms and descriptor displays (see Illustration E.) The Thesaurus also contains a "rotated descriptor display" (Illustration F) which allows access to the vocabulary from any word in the descriptor list. (Note: An edition of this rotated descriptor display, later than the 7th printed edition, (1977) also contains words in the lead-in terms, e.g., "Prediction" is in the display from the lead-in term "Achievement Prediction Use Achievement Tests").

Online, in LRS's DIALOG, where the ERIC Thesaurus is available, the descriptor displays (see Illustration F) are immediately accessible by a "double expand" command. Because this command accesses the Basic Index file for ERIC, postings are listed (under "Items") and for each term in the descriptor display (R1, R2, etc.) the type of term it is is explained (U for Use, N for Narrow, B for Broad, R for Related). Opposite each descriptor, the number of related terms (RT) each has is also given.

As for the rotated descriptor display, there is no equivalent online at the present time.

In fact, some "strange" lists occur if you search single words in the Basic Index of each vendor's version of ERIC.

If "Rating" is the word you are searching (as Illustration F shows it) in print, you would find four descriptors in the Rotated Descriptor display. Online, in LRS, DIALOG; for example, you would find some words above it in the Basic Index and below it. Occasionally you find mis-spelled words (for

Illustration E. ERIC Thesaurus--In Print and Online on LRS.

ALPHABETICAL DISPLAY

Lead-In Term

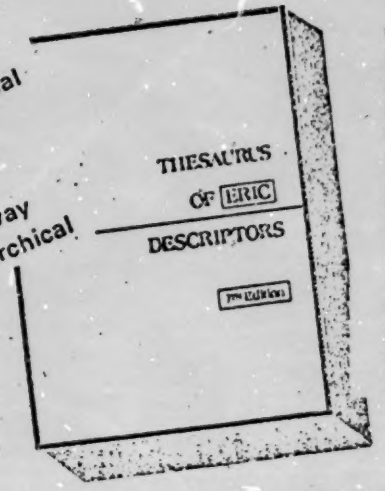
Achievement Prediction
USE ACHIEVEMENT TESTS

Descriptor

ACHIEVEMENT RATING Jul 1966
CIJE 184 RIE 267
SN The process of comparing achieved performance and the ranking assigned to compared performances
UF Achievement Comparison
NT Grades (Scholastic)
BT Evaluation
RT Academic Achievement
Academic Records
Achievement
Achievement Tests
Awards
Grading
Merit Rating Programs
Pass Fail Grading
Progressive Retardation
Rating Scales
Report Cards
Student Evaluation

ACHIEVEMENT TESTS Jul 1966
CIJE: 460 RIE: 1,127
UF Achievement Prediction
NT College Entrance Examinations

DISPLAYS:
Alphabetical
Rotated
Two-way Hierarchical

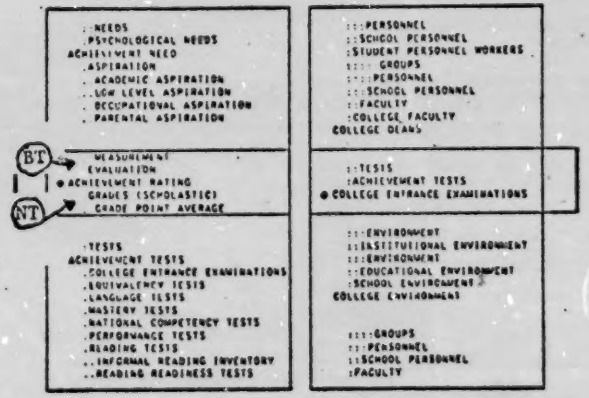


LRS

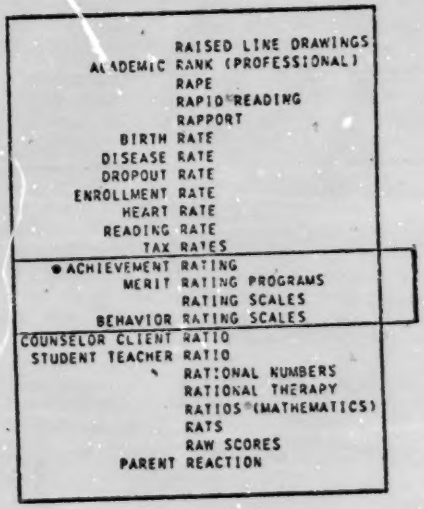
E (ACHIEVEMENT RATING)

REF	INDEX-TERM	TYPE	ITEMS	RT
P1	ACHIEVEMENT RATING (THE PROCESS OF COMPAR----		531	15
R2	ACHIEVEMENT COMPARISON--U			1
R3	GRADES (SCHOLASTIC)-----N		843	18
R4	EVALUATION----------B		55040	58
R5	ACADEMIC ACHIEVEMENT----R		9799	43
R6	ACADEMIC RECORDS--------R		160	7
R7	ACHIEVEMENT----------R		21858	28
R8	ACHIEVEMENT TESTS-------R		1966	18
R9	AWARDS----------R		797	7
R10	GRADING----------P		1868	10
R11	MERIT RATING PROGRAMS---R		67	8
R12	PASS FAIL GRADING-------R		181	8
R13	PROGRESSIVE RETARDATION-P		7	11
R14	RATING SCALES----------R		2125	8
R15	REPORT CARDS----------R		75	7
R16	STUDENT EVALUATION------P		4035	20

TWO-WAY HIERARCHICAL TERM DISPLAY



ROTATED DESCRIPTOR DISPLAY



Expand in LRS

LRS

```
? E Rating
REF  INDEX-TERM                    TYPE ITEMS RT
E1   RATIFIED                          19
E2   RATIFIERS                          1
E3   RATIFY                             2
E4   RATIFYING                          3
E5   RATILIN                            1
E6  -RATING                          5567
E7   RATING SCALE                       1
E8   RATING SCALE FOR
       CHILDREN (BILLER)                1
E9   RATING SCALE FOR
       CURRICULUM EVALUATION-           1
E10  RATING SCALE FOR
       DIRECTED TEACHING                1
E11  RATING SCALE FOR
       TEACHERS                         1
E12  RATING SCALES                   2128  8
E13  RATING SCALES                      1
E14  RATING/RANKING SCALE OF
       CHILD BEHAVIOR                   1
E15  RATINGS                         2699
                                    -MORE-
?E E12
REF INDEX-TERM                     TYPE ITEMS RT
P1   RATING SCALES                  2128  8
P2   BEHAVIOR RATING SCALES--N      869  6
P3   GRADE EQUIVALENT SCALES-N      26  7
P4   INTEREST SCALES----------N     150  4
P5   ACHIEVEMENT RATING-------R     531  15
P6   FORCED CHOICE TECHNIQUE-R      79  4
P7   MEASUREMENT-------------R   12005 .32
P8   Q SORT-----------------R       91  12
P9   SEMANTIC DIFFERENTIAL---R     252  5
?
```

In Print
Rotated Descriptor Display

```
                RAISED LINE DRAWINGS
      ACADEMIC RANK (PROFESSIONAL)
                RAPE
                RAPID READING
                RAPPORT
           BIRTH RATE
         DISEASE RATE
         DROPOUT RATE
      ENROLLMENT RATE
           HEART RATE
         READING RATE
             TAX RATES
  • ACHIEVEMENT RATING
          MERIT RATING PROGRAMS
                RATING SCALES
        BEHAVIOR RATING SCALES
 COUNSELOR CLIENT RATIO
  STUDENT TEACHER RATIO
                RATIONAL NUMBERS
                RATIONAL THERAPY
                RATIOS (MATHEMATICS)
                RATS
                RAW SCORES
          PARENT REACTION
```

Illustration F(1). In Print Rotated Description Display and Online Browsing in Basic Index on LRS, SDC and BRS.

NBR in SDC

SDC

```
YOU ARE NOW CONNECTED TO THE ERIC DATABASE.
COVERS RIE 1966 THRU JUN 1979 (7906) AND
CIJE 1969 THRU FEB 1979 (7902)

SS 1 /C?
USER:
NBR RATING

PROG:
POSTINGS     TERM
       1     *RATILIN/IT
       1     *RATILIN/IW
     281     RATING/TI
    5640     RATING/BI
       1     RATING/IT
UP N OR DOWN N?

USER:
DOWN 10

PROG:
POSTINGS     TERM
    8782     RATING/IW
    1793     *RATING/IW
       1     RATING SCALE/IT
       1     RATING SCLAE FOR CHILDREN (BILLER)/IT
       1     *RATING SCALE FOR CHILDREN (BILLER)/IT
       1     RATING SCALE FOR CURRICULUM EVALUATI/IT
       1     RATING SCALE FOR DIRECTED TEACHING/IT
       1     *RATING SCALE FOR DIRECTED TEACHING/IT
       1     RATING SCALE FOR TEACHERS/IT
    2179     RATING SCALES/IT
UP N OR DOWN N?

USER:
DOWN 10

PROG:
POSTINGS     TERM
     953     *RATING SCALES/IT
       1     RATING SCLAES/IT
       1     *RATING SCLAES/IT
       1     RATING/RANKING SCALE OF CHILD BEHAVI/IT
     482     RATINGS/TI
    2815     RATINGS/BI
      21     RATINGS/IW
       8     *RATINGS/IW
       1     RATINQ/BI
       1     RATINS/TI
UP N OR DOWN N?

USER:
DOWN 5

PROG:
POSTINGS     TERM
       1     RATINS/BI
      63     RATIO/TI
    1579     RATIO/BI
     437     RATIO/IW
     121     *RATIO/IW
```

Root in BRS

BRS

```
1_:    ROOT TRANSPORT
       TRANSPORT$
       TRANSPORT                    169 DOCUMENTS
       TRANSPORTABILITY              33 DOCUMENTS
       TRANSPORTABLE                 58 DOCUMENTS
       TRANSPORTATIN                  1 DOCUMENT
       TRANSPORTATION              2953 DOCUMENTS
       TRANSPORTATIONAL               1 DOCUMENT
       TRANSPORTATIONS                6 DOCUMENTS
       TRANSPORTED                   51 DOCUMENTS
       TRANSPORTEES                   1 DOCUMENT
       TRANSPORTERS                   1 DOCUMENT
       TRANSPORTING                  63 DOCUMENTS
       TRANSPORTION                   1 DOCUMENT
       TRANSPORTS                     6 DOCUMENTS
```

Illustration F(2). Browsing in Basic Index of ERIC Online on LRS, SDC and BRS

LRS

```
? E LIBRARIES
REF   INDEX-TERM              TYPE ITEMS RT
E1    LIBRARIAN/S ------------        2
E2    LIBRARIANSHIP ----------      580  1
E3    LIBRARIANSHIP...JUNE----        1
E4    LIBRARIANSHIP:
        UNIVERSITY OF OREGON--        1
E5    LIBRARIE ---------------        2
E6   -LIBRARIES --------------     8369 49
E7    LIBRARIES...AUG --------        1
E8    LIBRARIES...DECEMBER----        1
E9    LIBRARIES...OCCASION ---        1
E10   LIBRARO ----------------        1
E11   LIBRARY ----------------    12667
E12   LIBRARY ACCOUNTABILITY -        1
E13   LIBRARY ACQUISITION (PRO
        CESS OF ACQUIRING) ---      583  6
E14   LIBRARY ACQUISITION ----        1
E15   LIBRARY ADMINISTRATION -      292 13
E16   LIBRARY ADMINISTRATORS--        1
E17   LIBRARY AIDES ----------              1
E18   LIBRARY AIDS -----------
```

SDC

```
USER:
NBR LIBRARIES/TI

PROG:
POSTINGS      TERM
    327       LIBRARIANS/TI
    198       LIBRARIANSHIP/TI
   1580       LIBRARIES/TI
      1       LIBRARO/TI
   4096       LIBRARY/TI
UP N OR DOWN N?

USER:
NBR LIBRARIES/IT

PROG:
POSTINGS      TERM
    129       LIBRARIANSHIP/IT
      1       LIBRARIANSHIP: UNIVERSITY OF OREGON/IT
   1778       LIBRARIES/IT
      1       LIBRARY ACCOUNTABILITY/IT
    583       LIBRARY ACQUISITION/IT
UP N OR DOWN N?
```

```
? S LIBRARIES/DF
          9  1777 LIBRARIES/DF
REF  INDEX-TERM              TYPE ITEMS RT
R1   LIBRARIES ---------------     8369 49
R2   BRANCH LIBRARIES -------N      40  5
R3   COLLEGE LIBRARIES ------N     825  8
R4   DEPOSITORY LIBRARIES----N      52  8
R5   FILM LIBRARIES ---------N     149 10
R6   NATIONAL LIBRARIES -----N     236  4
R7   PUBLIC LIBRARIES -------N    1533  8
R8   REGIONAL LIBRARIES -----N     168  8
R9   RESEARCH LIBRARIES -----N     286  6
R10  SCHOOL LIBRARIES -------N    1004  5
R11  SPECIAL LIBRARIES ------N     611  9
R12  STATE LIBRARIES --------N     325  4
R13  UNIVERSITY LIBRARIES ---N    1555  5
R14  FACILITIES -------------B   11760 82
R15  ARCHIVES ---------------R     441 11
R16  BOOK CATALOGS ----------R      94  8
R17  CATALOGING -------------R
```

BRS

```
1_:   ROOT LIBRARIES                5763 DOCUMENTS
      LIBRARIES$                       1 DOCUMENT
      LIBRARIES                        7 DOCUMENTS
      LIBRARIES.DECEMBER               1 DOCUMENT
      LIBRARIES--A                     1 DOCUMENT
      LIBRARIES--ACADEMIC              1 DOCUMENT
      LIBRARIES--ACQUISITION           1 DOCUMENT
      LIBRARIES--AN                    1 DOCUMENT
      LIBRARIES--ARE                   1 DOCUMENT
      LIBRARIES--BIBLIOGRAPHY          1 DOCUMENT
      LIBRARIES--HOUSTON               1 DOCUMENT
      LIBRARIES--HOW                   1 DOCUEMNT
      LIBRARIES--INCLUDING             1 DOCUMENT
      LIBRARIES--LAS.                  1 DOCUMENT
      LIBRARIES--LIBRARY               1 DOCUMENT
      LIBRARIES--LIKES                 1 DOCUMENT
      LIBRARIES--PHILOSOPHY            2 DOCUMENTS
      LIBRARIES--PUBLIC                1 DOCUMENT
      LIBRARIES--SCHOOL                4 DOCUMENTS
      LIBRARIES--SPECIFICALLY          1 DOCUMENT
      LIBRARIES--THE                   1 DOCUMENT
      LIBRARIES--TO                    1 DOCUMENT
      LIBRARIES--TWO-YEAR              1 DOCUMENT
      LIBRARIES--WERE                  1 DOCUMENT
      LIBRARIES--WHY                   1 DOCUMENT
      LIBRARIES--YES                   2 DOCUMENTS
      LIBRARIES--38                    4 DOCUMENTS
      LIBRARIES-A                      1 DOCUMENT
      LIBRARIES/MEDIA
      LIBRARIES/SCHOOL
```

example, Rating Scales); "rating scale" as the first words in the titles of several items, and so on, but you do not find the other three descriptors with rating embedded in them -- even if you "double expand" the descriptor rating scales to expose its display because these other descriptors are not "related" to it.

If online you send in a lead-in term, you will not see the "Use" reference. However, as a fluke, you may see that a lead-in term (e.g., Entrance Examinations) has a posting because some indexer used it as an identifier!

```
S  ENTRANCE EXAMINATIONS
             4      1 ENTRANCE EXAMINATIONS
?  T4/2/1
4/2/1
EJ007763  VT500705
   CAREERS ORIENTATED EDUCATION
   O'NEILL, HUGH
   TECH EDUC IND TRAINING, 11, 8, 323-324, 328     69 AUG
   DESCRIPTORS:  ADULT EDUCATION/ *CAREER OPPORTUNITIES/ *CURRICULUM/
*OCCUPATIONAL GUIDANCE/ VOCATIONAL EDUCATION
   IDENTIFIERS:  *ENGLAND/ ENTRANCE EXAMINATIONS

?
```

Each retrieval system will show you a different array in the Basic Index, depending on what fields they have chosen the words from and how they haye handled punctuation in those fields. (See Illustration F(1) and F(2) for BRS and SDC.)

SDC labels each line of basic index by field from which the term came (TI, IW, etc.) repeating the line if it were used as a major descriptor or major identifier (*/IW).

BRS does not remove punctuation from title field (--). This explains unusual (and useless?) listings when you browse in BRS using ROOT command. The single hyphen (-) may come from title, descriptor and identifier field.

D. Recommendations for the Online ERIC Record

1. There should be some uniformity in the codes used on all commercial retrieval systems to designate the data elements in the ERIC record.

   Central ERIC should work with ANSI, NFAIS, ASIDIC and other coordinating bodies to bring this about, across retrieval systems and across data bases.

   The non-uniform codes are not many (see Figure 1), but the differences can cause confusion when prints from online and offline searches are examined along with the printed versions of RIE and CIJE.

2. The order of printing the data elements in the ERIC record should be as uniform as possible across all retrieval systems. As a corollary, when new fields are added, some of which are not printed in RIE or CIJE, Central ERIC should specify to retrieval vendors where these should appear in online/offline prints. Case in point are the new fields in the ERIC Record (GOV, LAN, GEO, and PT).

3. The order of printing items from the ERIC file should be uniform--in printed indexes and online. Here, I think, the online versions (reverse order of ED or EJ order) is preferable to order by first word in title of item.

4. When the online/offline output is a mix of ED and EJ numbers, the output of all systems should be uniform, either all ED first, then EJ or all ED and EJ from current year (so designated).
   Sorting options for offline prints can customize the output, of course, but the default condition should be uniform.

E. Specific Recommendations for the ERIC Online Basic Index and Online Thesauurus

1. The fields from the ERIC record used to form the Basic Index of single word terms should be uniform across the three commercial retrieval systems, LRS, SDC, and BRS.

2. The treatment of ERIC descriptors and identifiers in such Basic Indexes should be uniform across all systems (preferably the most liberal treatment possible without creating confusing in postings information, as at present on DIALOG, or as restrictive, as at present on BRS).

3. The online ERIC Thesaurus should have at least as many features as the printed version, e.g., rotated descriptor display, two way hierarchy, descriptor group listings, lead-in vocabulary access to displays.

4. Preferably the online basic indexes and thesaurus should provide more access to the ERIC vocabulary and improve the search retrieval results than at present. Other portions of this report discuss what these improvements might be.

At this point, let us paraphrase what Aitchison and Gilchrist* said in 1972:

A thesaurus need not be thought of as synonymous with "controlled vocabulary." The choice in thesaurus construction need no longer be either a "free language" or a "controlled thesaurus. We suggest searching thesauri to ensure matches. Retrieval systems provide truncation and word proximity features which create, momentarily, evidence of the vocabulary terms and phrases in the data base. These records should be available for longer than the moment in online versions of the "rotated descriptor displays," KWIC indexes to titles of items, etc. Hierarchical displays of broader-narrower terms could be displayed online with both descriptor postings and "free text" items.

A free language thesaurus is similar in structure to the controlled language thesaurus. It should provide synonyms, quasi-synonyms, alternative spellings and abbreviations and all possible alternative word forms. It must deal with multi-word concepts, including them in the system either as main concepts or entry terms. Hierarchical and other relationships between terms are very necessary, and these may be shown both alphabetically and systematically.

A free language thesaurus differs from a controlled language thesaurus in that:

a. It is not used in indexing.
b. The form of the main concepts in the thesaurus need not be controlled rigidly, so long as alternative forms are shown as entry terms.

*Thesaurus Construction; A Practical Manual. Aslib, 1972.

c. The problem of accommodating new terms is not so urgent as in the controlled language thesaurus, when a speedy decision has to be made by the indexers, so as not to delay input.
d. The purpose of the thesaurus is to save the searcher the intellectual effort of recalling possible terms and term relationships. The thesaurus need not be the sole index language available to the searcher-other published thesauri, classification systems, dictionaries, encyclopedias may also be included.

Suggestions later in this part of the report show what an "online free language thesaurus" might look like and what it might do to aid the online searcher. It is not conceived as an indexer's thesaurus alone and would never appear on print. Such a recommendation, to be implemented, would require some very carefully planned, coordinated research between ERIC and the Lockheed Retrieval System staff who have the design responsibilities for the only commercial version of an online ERIC Thesaurus. Reference to COSMOS, SMERC's system might be helpful.

II. HOW IS ERIC SEARCHED ONLINE?

A. Background

Every retrieval system vendor has training sessions and user materials (user guides, manuals, online aids) for their system and for every data base on their system. Some of them have even created special online training films (e.g., File 201 on DIALOG--ERIC ONTAP).

Every retrieval system prides itself on its competitive edge because it has features (for searching, printing, saving searches, online searching aids, etc) no other system has. Every system adds or revises its command language/ features to mee popular demands. Witness "Superselect" on DIALOG, ORBIT IV on SDC, etc.

These different features may be unique and novel, or unique and archiac. Over the years, the systems have tended to look more and more similar in function but not in command language. Writers like Martin, Atherton, Marcus, Lawrence and Bearman* and others have addressed this topic and the problems and possibilities it poses for online searchers.

In other parts of this report we have itemized the differences in searching ERIC online. We have also summarized findings from experienced online searchers and from our own retrieval tests.

---

*Martin, T.H., A Feature Analysis of Interactive Retrieval System, Stanford University, Institute for Communications Research. Sept. 1974. (SU-COMM-ICR-74-1)

Atherton, P. "Standards for a User-System Interface Language in On-Line Retrieval Systems," Online Review 2 (1978), 57-61.

Marcus, R.S., Computer Interfaces for User Access to Heterogenous Information Retrieval Systems. M.I.T. Dept. of Electrical Engineering and Computer Science. Electronic Systems Laboratory. April 1977. (ESL-R-739).

Lawrence, B. and T.C. Bearman, On-Line Commands; A User's Quick Guide for Bibliographic Retrieval Systems. National Federation of Abstracting and Indexing Services. March, 1977.

As Lancaster and others* have indicated, the topic is a complex one because it involves a study of several variables:

- the retrieval system's capabilities
- the data base as it is online on the system
- the vocabulary as it is online on the system
- the online searcher's use of all of the above
- the end user for whom the search is performed
- the end user's query in his/her own mind
- the end user's search objective
- the online searcher-end user interface

It is not our intention in this report to cover this topic in any detail, but we have displayed variations in searching/printing capabilities of the three major retrieval system vendors because this cross-system comparison is not often done in a way specific to a single data base. But more of this kind of comparison is necessary to aid the online searcher who uses more than one system.

An unpublished 1975 report by Bonnie Gratch, then a student at Syracuse University, documented a comparative analysis of operating features of BRS (then SUNY-BCN), LRS/DIALOG and SDC/ORBIT as they affected the ERIC database. She included only selected features to highlight the state-of-the-art when compared with a proposed list of prerequisite characteristics of a good interactive bibliographic retrieval system.

Up to this time (from April 1975 on) no system provided these "desirable" characteristics:

1. Monitoring use online, if requested, to help with problems.
2. System compensation for common errors in spacing, mis-spelling, or punctuation.
3. Ranking or output.
4. Hierarchical searching of ERIC descriptors.
5. Automatic access to indexing records from free text matches or lead-in vocabulary hits.

---

*Lancaster, F.W., The Measurement and Evaluation of Library Services, Information Resources Press, 1977. pp. 378-379.

Rouse, S.H. and L.W. Lannon, "Some Differences Between Three Online Systems," Online Review 1 (June 1977), 117-132.

Burrows, S. and S. Kyle, "Searching the MEDLARS File on NLM and BRS," Bulletin of the Medical Library Association 67 (January 1979), 15-24.

6. Matrix display of descriptor/free text hits for search/indexing comparisons.
7. History file for obsolete, discarded terms, with dates of use.

Some of these features would, no doubt, help the ERIC online searcher.

To augment this wish-list, we turned to Marcia Bates, of the University of Washington Library School who has been studying the search tactics of information system users. It is her opinion, and we concur, that retrieval systems should be able to do more than they do to help the user online when she/he is involved in the search process. Offline training aids aside, the needs of the searchers after they have signed on, must be answered if we are ever to make online bibliographic searching low-cost, cost-effective, and efficient.

Dr. Bates' focus was different from any others we had seen. She started with the problems searchers face online. For example:

... TOO MANY DOCUMENTS
... TOO FEW DOCUMENTS
... NO DOCUMENTS
... NEED OTHER TERMS
... NEED TO REVISE TERMS

She then analyzed what they do themselves faced with such a problem. For example, she has itemized such tactics as PINPOINT, EXHAUST, BLOCK, REDUCE, RESPACE, RESPELL, RELATE, FIX, etc.

Next she asked: what of these user responses could be automated, i.e., built-into the retrieval system itself? Some of these tactics or responses are already features in the systems, such as NEIGHBOR, TRUNCATE, TRACE, but many are not.

We will not report her entire report here, but it might prove useful to report here her "Summary of Information Search Tactics and Definitions"

(Bates-Table 1). We then display how she thinks these tactics could be regrouped as automatic "System Responses" when one of the five searcher's problems surfaces. (Bates-Table 2)

She foresees at least four levels of use of these new system features--with more and more automatic system response built in, as possible, or as required or requested by the user.

To quote from her report to us, here is one invented example which would show how the searcher would be aided online by the capability to execute tactics built into the system, at the user's request:

> Suppose that with a given problem the searcher anticipates a large complex search. There are many concepts to be combined and many of the concepts are likely to have synonyms and spacing or spelling variants. The searcher anticipates having to tinker a lot, dropping and adding parallel terms (variants of a single concept), as well as dropping and adding whole concepts (with or without variants).
>
> Herewith is an invented example that is simpler than many actual problems searchers encounter daily: The problem is to find out about the training of elementary school teachers in mathematics. With the proposed approach the searcher generates immediately a large pool of variant terms through the use of NEIGHBOR and the like. Then the entire search, with all concepts and all variant terms for each of the concepts, with the help of Boolean operators and parentheses, is input at the beginning of the search into a special designated storage cell. In the example above, this "maxi-search" might be something like:
>
> ((Training OR Instruction) AND (Elementary school OR Primary school OR K-6) AND (Teachers OR Instructors) AND Mathematics)
>
> This stored maxi-search then becomes the basis for all or a good part of the rest of the search on the topic. Through the use of tactics, the searcher can try in quick succession a large variety of combinations of the terms in the maxi-search. From then on in the search, terms, for the most part, do not have to be repeated, and Boolean combinations of terms and concepts can be carried out by the input of brief commands, standing for various tactics. It is possible for the searcher to try many variant combinations of the terms within the maxi-search in rapid order, without the usual worry about typing errors in strings of 20 or 30 or 50 characters, or of forgetting to try one or more of the logical combinations of terms.

Bates-Table 1. Summary of Information Search Tactics and Definitions

MONITORING TACTICS	Tactics to keep the search on track and efficient.
M1. CHECK	To review the original request and compare it to the current search topic to see that it is the same.
M2. WEIGH	To make a cost-benefit assessment, at one or more points of the search, of current or anticipated actions.
M3. PATTERN	To make oneself aware of a search pattern, examine it, and redesign it if not maximally efficient or if out of date.
M4. CORRECT	To watch for and correct spelling and factual errors in one's search topic.
M5. RECORD	To keep track of trails one has followed and of desirable trails not followed up or uncompleted.
FILE STRUCTURE TACTICS	Techniques for threading one's way through the file structure of the information facility to desired file, source, or information within source.
F1. BIBBLE	To look for a bibliography already prepared, before launching oneself into the effort of preparing one; more generally, to check to see of the search work one plans has already been done in a usable form by someone else.
F2. SELECT	To break complex search queries down into sub-problems and work on one problem at a time.
F3. SURVEY	To review, at each decision point of the search, the available options before selecting.
F4. CUT	When selecting among several ways to search a given query, to choose the option that cuts out, eliminates, the largest part of the search domain at once.
F5. STRETCH	To use a source for other than its intended purposes.
F6. SCAFFOLD	To design an auxiliary, indirect route through the information files and resources to reach the desired information.
F7. CLEAVE	To employ binary searching in locating an item in an ordered file.

Bates-Table 1 (con't.)

SEARCH FORMULATION TACTICS	Tactics to aid in the process of designing or re-designing the search formulations.
S1. SPECIFY	To search on terms that are as specific as the information desired.
S2. EXHAUST	To include most or all elements of the query in the initial search formulation; to add one or more of the query elements to an already-prepared search formulation.
S3. REDUCE	To minimize the number of elements of the query in the initial search formulation; to subtract one or more of the query elements from an already-prepared search formulation.
S4. PARALLEL	To make the search formulation broad (or broader) by including synonyms or otherwise conceptually parallel terms.
S5. PINPOINT	To make the search formulation precise by minimizing (or reducing) the number of parallel terms, retaining the more perfectly descriptive terms.
S6. BLOCK	To reject, in the search formulation, items containing, or indexed by, certain term(s), even if it means losing some document sections of relevance.
TERM TACTICS	Tactics to aid in the selection and revision of specific terms within the search formulation.
T1. SUPER	To move upward hierarchically, to a broader (superordinate) term.
T2. SUB	To move downward hierarchically, to a more specific (subordinate) term.
T3. RELATE	To move sideways hierarchically, to a coordinate term.
T4. NEIGHBOR	To seek additional search terms by looking at neighboring terms, whether proximate alphabetically, by subject similarity, or otherwise.
T5. TRACE	To examine information already found in the search in order to find additional terms to be used in furthering the search.
T6. VARY	To alter or substitute one's search terms in any of several ways.

00 142

Bates-Table 1 (con't.)

T7.	FIX	To try alternative affixes, whether prefixes, suffixes or infixes.
T8.	REVERSE	To reverse or rearrange the words in search terms in any or all reasonable orders.
T9.	CONTRARY	To search for the term logically opposite from that describing the desired information.
T10.	RESPELL	To search under a different spelling.
T11.	RESPACE	To try spacing variants.
IDEA TACTICS		Tactics to help generate new ideas or solutions or problems in information searching.
I1.	THINK	To stop and think about the search and try to come up with new ideas for solving search difficulties.
I2.	BRAINSTORM	To generate many ideas and to suspend critical reactions until the ideas are well-formed and can be fully evaluated.
I3.	MEDITATE	To develop the skill of achieving quickly a mind state in which thinking and intuitive capacities work together in solving search problems.
I4.	CONSULT	To ask a colleague for suggestions or information in dealing with a search.
I5.	RESCUE	To check for possible productive paths still untried, in an otherwise unproductive approach.
I6.	NOTICE	To watch for clues that revise one's notions of the nature of the question or of the answering information.
I7.	WANDER	To move among one's resources, being receptive to alternative sources and new search ideas triggered by the materials that come into view.
I8.	CATCH	To catch oneself in an unproductive search, to notice that a change in approach might be more productive.
I9.	BREAK	To break an habitual search pattern, that is, put it aside temporarily, in order to take a search tack more suited to the particular problem in question.

I10. BREACH	To breach the boundaries of one's region of search, to revise one's concept of the limits of the intellectual or physical territory in which one searches to respond to a query.
I11. JOLT	To think "laterally" (after Edward de Bono's concept, 10) to snap out of conventional thought patterns, and redefine a problem in a dramatically different way.
I12. STOP	To cease searching temporarily on a problem and do something else instead.
I13. CHANGE	To change something, anything, in one's search behavior--to try a different source, a different term, a different subject field, etc.
I14. RETRY	To search in a source again, after an initial failure to find in it the desired information.
I15.FOCUS	To look at the query more narrowly, in one or both of two senses: 1) to move from the whole query to a part of it, 2) to move from a broader to a narrower conceptualization of the query.
I16. DILATE	To look at the query more broadly, in one or both of two senses: 1) to move from a part of the query to the whole query, 2) to move from a narrower to a broad conceptualization of the query.
I17. SKIP	To shift laterally in one's view of the query, in one or both of two senses: 1) to move from searching one part of a complex, multi-part query to another one (Cf. F2, SELECT), 2) to view the query from, and search in, another angle, that angle being neither broader or narrower, but simply different.
QUESTION-NEGOTIATION TACTICS	At this point these tactics are proposed and not developed--except for one tactic useful as an idea tactic, listed below:
Q1. REFRAME	To examine implicit frame-of-reference information coming along with the user's query, deal with it explicitly with the user, and negotiate changes in the query to eliminate distortions.

27

Bates Table 2. Suggested tactics in response to requests.

SEARCHER PROBLEM	SYSTEM RESPONSE LIST	DEFINITION*
"TOO MANY DOCUMENTS"	SPECIFY	S1, S2
	EXHAUST	
	PINPOINT	
	BLOCK	
	SUB	T2
"TOO FEW DOCUMENTS"	NEIGHBOR*	T4, T5
	TRACE*	S3, S4
	REDUCE	
	PARALLEL	
	FIX*	T7, T1
	SUPER	
	RELATE	T3, T6
	VARY	
"NO DOCUMENTS"	RESPACE	T11, T10
	RESPELL	
	REVERSE	T8, T9
	CONTRARY	
	SUPER	T1, T3
	RELATE	
	NEIGHBOR	T4, T8
	TRACE	
"NEED OTHER TERMS"	NEIGHBOR	T4, T5
	TRACE	
	SUPER	T1, T2
	SUB	
	RELATE	T3
"NEED TO REVISE TERMS"	RESPACE	T11, T10
	RESPELL	
	FIX	T7, T8
	REVERSE	
	CONTRARY	T9
	SUPER	
	SUB	T1, T2
	RELATE	T3

*See Bates Table 1.

This approach can be illustrated through the example above. Once the maxi-search has been input, the searcher may key in, for example, "E, PI", to stand for EXHAUST, PINPOINT. In this case, the system would then search on all four concepts (EXHAUST), but only on the first term in any OR'd set (PINPOINT). (It is assumed that in inputting the maxi-search the searcher has taken the precaution of inputting the "best", that is, the most precise, terms first in each OR'd sequence of parallel terms for the concepts.) This search on the most precise terms is just the meaning of PINPOINT. Thus, in inputting E,PI (EXHAUST, PINPOINT), the searcher has gotten a search on the following terms in the example of maxi-search: Training AND Elementary School (EXHAUST minus 1, PINPOINT), would produce even more--in our example, a succession of searches on each combination of three concepts within the four most precise in the maxi-search. These would be typed out in full by the system with associated postings, e.g.,

Training AND Elementary School AND Teachers	42
Training AND Elementary School AND Mathematics	12
Elementary School AND Teachers AND Mathematics	18
Training AND Teachers AND Mathematics	51

"E-2" would lead to searches on all combinations of two of the four terms, etc. If the searcher preferred, the same commands might be represented by R+ (REDUCE plus) commands instead.

Further complications could be introduced by varying the number of OR'd variants of each concept through the use of PINPOINT or PARALLEL. For example, "E-1, PI+1" would produce all combinations of the concepts in the maxi-search minus one (as above) and within each concept the first variant will be OR'd to the second, but no subsequent, variants. Using the example above, "E-1, PI+1" would produce this search: (Training OR Instruction) AND (Elementary school OR Primary School AND (Teachers OR Instructors) AND Mathematics. In this case, the only term left out is "K-6", one would have to include "PI+2", or some equivalent command in the search statement.

BLOCKed terms could be included in the maxi-search as well and included or withdrawn at will. For example, the maxi-search might have included "NOT Computer-aided instruction". A command of "E, PI, B" would tell the system, in addition to the EXHAUST, PINPOINT commands, to BLOCK "Computer-aided instruction". Omission of the B in later commands would allow CAI to be included in answer documents. Or the same effect might be achieved in the reverse manner, with NOT'd terms automatically BLOCKed unless summoned by some direct command such as "INCLUDE" (not currently in the tactics set).

Thus, through the use of tactics on command the searcher could easily search a very large number of the logical combinations implicit in the maxi-search, without having to retype terms each time or remember whether all logical combinations had been tried. A master Command should also be available, something like COMBINE (another term name would have to be found

for DIALOG), which would automatically search on all EXHAUST/REDUCE/PIN-POINT/PARALLEL/BLOCK combinations of the maxi-search and print them out with postings. The searcher could then scan the combinations and select promising ones to have some citations printed out. In many cases this master command would produce too many combinations, including nonsensical ones, and the searcher would prefer the more specific tactics of narrower scope, discussed above.

In summary, the idea here is that the searcher may input all anticipated elements of the search, as generated perhaps through a quick and dirty search using NEIGHBOR, TRACE, and similar tactics, and then to have brief commands available that produce a large variety of the logical combinations implicit in the original maxi-search. The searcher is then saved the necessity of laboriously typing out all manner of logical combinations. Note that this requires no "intelligence" on the part of the system, no sophisticated programs to detect variant terms, or the like. All the terms are put in at the beginning by the searcher. Subsequent brief commands require just the sort of mechanical manipulations of combinations that the computer does best.

If, in addition, the data base being searched contained information on hierarchical relationships, then it would be possible to expand the search beyond the terms contained in the maxi-search. Commands of SUPER, SUB, or RELATE could either produce searches directly on the related terms, or else the terms could be printed out by the system and the searcher designate which should be searched on.

B. Recommendations for Searching ERIC Online:

1. Some ERIC committee (or designated Clearinghouse) should study differences in ERIC retrieval results across systems.
2. This ERIC committee or CERIC should suggest new system features to improve ERIC online searching (see Bates work and other parts of this report.)
3. This ERIC committee should work with retrieval system vendors to improve their system-specific online ERIC user materials.
4. This ERIC committee should keep a running record/bibliography of ERIC online searching aids produced anywhere and periodically produce and distribute this bibliography to online user groups or have it published in journals such as Online or Database.
5. Within the ERIC "family", online searching aids and retrieval reports should be reviewed and approved by this ERIC committee. Such reports would include the recent PS bibliography/IAP on "Attachment and Children" by Linda Baskin or the ERIC/IR Markey-Atherton ERIC/ONTAP publication or the ERIC/IR Kabby Mitchell publication on computerized literature searching. (see bibliography of ERIC Searching Aids in the Introduction of this Project Report.)
6. Various features for limiting searches exist online at the present time (e.g., update, year of publication, document type, Major-Minor descriptions, etc.). A careful study of the differences across systems should be investigated.

III. WHAT ABOUT THE ERIC ONLINE USER?

A. Background

The third issue (Nov. 1978) of the Directory of ERIC Computerized Search Services contains some 341 entries from 49 of the 50 states in the USA. We know this to be the tip of the iceberg in identifying individual online ERIC users.

The 1974 Wanger study, the Impact of Online Services and now a recent LINK study ("Strategies in the Online Database Marketplace") list ERIC among the top four most frequently used databases. Martha Williams has estimated over one million searches are made each year online in such bibliographic data bases as ERIC. The availability of ERIC Tapes on all three major commercial retrieval systems and many other state-wide or academic computer-based retrieval services and on SMERC assures us that a great many people (possibly thousands) are searching ERIC online every day all over the world.

These ERIC online users use different computer terminal equipment and different retrieval systems. They serve different clientele at different intervals. They receive different kinds of training and refresher training. They use one or many systems, one or many databases. They charge differently for their services and they prepare differently for their online sessions. They work online with different constraints and capabilities.

With so many differences, how could we hope to present a composite picture of online ERIC searches? We really couldn't even begin to try because the data just do not exist, nor do we know of any reliable means for tapping the data sources. Directories are incomplete; online search profiles are not available or are confidential.

Nonetheless, we used what resources existed during the six months of this special project (September, 1978 to March 1979). We tapped a few people on their shoulder and talked to them. We heard about their online problems and practices, about their wish-lists and their habits. These conversations, in summary, are reported in another part of this report.

When we realized that some data on practices regarding controlled versus free text searching might not be forthcoming, we decided to run some experimental tests, building on a previous ERIC/IR publication by Markey and Atherton (ED 160 108). This study forms another part of this report.

Over 1/3 of the libraries surveyed in the recent LINK study (N = 316) said they plan to, or are considering, the cancellation of printed product subscription because they have adequate access to information online. This poses some real problems for data base suppliers and printed product users. ERIC is immune to many of these trends, but another, related problem needs to be noted.

In a guest editorial in DATABASE (March, 1979, p. 4-5), Sarah T. Kadec, the Assistant Director of the Information Management and Service Division, Executive Office of the President, in the White House, asked several thought-provoking questions about the need for effective management of the vast array of online databases. She begins by questioning the practicality and cost effectiveness of the present modus operandi which, in most libraries and information centers, is to acquire access to almost any database available which might have information in a subject area of concern to its users. Duplication, use, staff training, and access of many databases are among those questions she lists. Her last question is very appropriate to our present concerns:

> "Should the access be provided by the database producer or is it better and even necessary to acquire access through a vendor?" (underlining added).

She feels that "We can no longer continue in the uncontrolled 'take-everything' atmosphere so prevalent today!" Sooner or later data base producers like ERIC will have to try to solve this user problem of inefficiency, redundancy and access. The issue for ERIC is aggravated by the present different image of ERIC on the three systems.

B. General Recommendations

1. Different search objectives of the end-user whom the online searcher is serving appears to present another dilemma for data base suppliers like ERIC.
2. To help the person who usually "wants a lot" (high recall), the free text portions of the ERIC record have to be more carefully prepared.
3. To access information on newer topics another vocabulary aid besides the present Thesaurus or the Identifier Authority Lists may need to be prepared.
4. Non-subject data elements in the ERIC record may need to be explained to users and tested for their ability to limit excessive output.

The pubdate and pubtype fields now in ERIC are not available for the entire ERIC database and are accessed differently in the different retrieval systems. This confuses the online user and a project should clear up this confusion.

Some online users have even more unusual uses of ERIC online. They request sub-sets of ERIC and other files, transfer these to their own private computer system and provide themselves with a customized ERIC-like database. WEECN, the Women's Educational Equity Communications Network, is a case in point, which has combined several files, re-indexed them, and produced a new abstract journal. Records from all other data bases have been reformatted and re-indexed using ERIC Thesaurus (augmented) so that

it all looks like ERIC (see Figure 4). This type of "tertiary" use of ERIC files may pose some interesting possibilities for combining all data bases for online searching. ERIC needs to consider the implications of such developments.

Figure 4. Psychological Abstracts Record Reformatted for Resources in Women's Educational Equity (Journal) Using ERIC Record Format and ERIC Thesaurus

Psychological Abstracts on LRS:

DOC YEAR: 1978 VOL NO: 59 ABSTRACT NO: 03453
MOTIVE TO AVOID SUCCESS, MOTIVE TO AVOID FAILURE, STATE-TRAIT ANXIETY, AND PERFORMANCE.
PATTY, ROSEMARIE A.; SAFFORD, SHARON F.
WAKE FOREST U
IN C.D. SPIELBERGER & I.G. SARASON (EDS), STRESS AND ANXIETY: IV. WASHINGTON, DC: HEMISPHERE, 1977. XVI, $18.95, 77-93
LANGUAGE: ENGL CLASSIFICATION: 31, 00
ADMINISTERED THE STATE-TRAIT ANXIETY INVENTORY (STAI), THE ACHIEVEMENT ANXIETY TEST, AND THE DIGIT SPAN SUBSCALE OF THE WAIS TO 60 UNDERGRADUATE FEMALES AND 60 UNDERGRADUATE MALES TO INVESTIGATE THE EFFECT OF TASK DIFFICULTY AND AMBIVALENT ACHIEVEMENT MOTIVES ON STATE ANXIETY AND PERFORMANCE. THE INTERACTION OF TASK DIFFICULTY AND AMBIVALENT MOTIVES PRODUCED CHANGES IN STATE ANXIETY AND PERFORMANCE AMONG FEMALES, AND IN STATE ANXIETY BUT NOT IN PERFORMANCE AMONG MALES. FOR FEMALES, THE RELEVANT ACHIEVEMENT MOTIVE WAS THE MOTIVE TO AVOID SUCCESS (MAS) AND, FOR MALES, THE MOTIVE TO AVOID FAILURE (MAF). INTERCORRELATIONS BETWEEN THE MAS, MAF, FACILITATING TEST ANXIETY, AND TRAIT ANXIETY RESULTED IN SIGNIFICANT CORRELATIONS BETWEEN MAF AND FACILITATING TEST ANXIETY FOR BOTH MALES AND FEMALES, AND BETWEEN TRAIT ANXIETY AND BOTH MAF AND FACILITATING TEST ANXIETY FOR FEMALES. (20 REF)
SUBJECT TERMS: HUMAN SEX DIFFERENCES, FEAR OF SUCCESS, TEST ANXIETY, ANXIETY, TASK COMPLEXITY, PERFORMANCE, ACHIEVEMENT MOTIVATION; 23510, 19424, 52190, 03310, 51480, 37525, 00500
INDEX PHRASE: TASK DIFFICULTY & AMBIVALENT ACHIEVEMENT MOTIVES, STATE ANXIETY & PERFORMANCE, MALE VS FEMALE COLLEGE STUDENTS

Reformatted and Reindexed according to ERIC for WEECN Data Base:

WEECN Accession Number—identification number sequentially assigned to documents as they are processed. — WE 000 000

Author(s) — *Patty, Rosemarie A. Safford, Sharon F.*

Title — Motive to Avoid Success, Motive to Avoid Failure, State-Trait Anxiety, and Performance

Author's Institutional Affiliation — Affil—Wake Forest Univ., Winston Salem, N.C.

Publication Date — Pub Date—77

Journal Citation — Journal Cit—Journal of Social Psychology; v101 n1 p59-67 February 1977

Pagination — Note—16p.; In "Stress and Anxiety: IV," C.D. Spielberger, Ed. and I. G. Sarason, p77-93

Descriptive Note

Ordering Information — Available from—Hemisphere Publishing Corp., 1025 Vermont Ave. N.W., Washington, D. C. 20005

Announcement Journal of source data base — Data Base ID—PAFEB78; 59-3453

Psychological Abstracts Accession Number

Descriptors (Only terms preceded by an asterisk are printed in the subject index.) — Descriptors—*Anxiety, Motivation, *Performance Factors, *Sex Differences

Identifiers—*Fear of Success

Identifiers (Terms not found in the ERIC Thesaurus. Only terms preceded by an asterisk are printed in the subject index.)

Abstract — The State-Trait Anxiety Inventory (STAI), the Achievement Anxiety Test, and the Digit Span subscale of the WAIS were administered to 60 undergraduate females and 60 undergraduate males to investigate the effect of task difficulty and ambivalent achievement motives on state anxiety and performance. The interaction of task difficulty and ambivalent motives produced changes in state anxiety and performance among females, and in state anxiety but not in performance among males. For females, the relevant achievement motive was the motive to avoid success (MAS) and, for males, the motive to avoid failure (MAF). Intercorrelations between the MAS, MAF, facilitating test anxiety, and trait anxiety resulted in significant correlations between MAF and facilitating test anxiety for both males and females, and between trait anxiety and both MAF and facilitating test anxiety for females. (20 ref)

Following are specific recommendations from the experienced online ERIC searchers we talked with and our own project staff's recommendations for improving online searching. We hope these recommendations will be read and studied carefully by Central ERIC, the ERIC Facility staff, and the staffs of the online retrieval system designers who are "part of" the expanded ERIC Information System.

C. Specific Recommendations from Online Searchers

We received comments like the following in conversations with ERIC online searchers when we asked them to: "Build the optimum retrieval system for searching ERIC online."

--- Let me be able to select terms and combine them in a single search statement with truncation and field delimiters. [BRS]

BRS allows the searcher to enter a combination of boolean and/or positional operators. Enclosing portions of a statement makes relationships clear and frees the searcher from keeping in mind the order in which the system treats logical and positional operators. For example:

(ONE TWO THREE) ADJ YEAR$1 ADJ OLD.TI.AB.ID.

As soon as BRS unbinds descriptors and enters them into the basic index, the logic used in the above statement can be extended across controlled vocabulary fields.

* * *

--- Change the system overflow message so that I can resume from the overflow. [BRS]

BRS overloads when I use truncation. For example, the system could do as follows: [SDC]

1___: EDUCATIONAL-A$

E1417 TOO MANY WORDS FOR EDUCATIONAL-A. RESULT 93475) INTERRUPTED (EDUCATIONAL-OBJECTIVES). CONTINUE? Y/N?

In this way, the searcher knows when the system ended processing and number of postings, and he/she can decide whether to continue the search or abort.

SDC offers this option for time overflow and stringsearching.

* * *

---

*Retrieval System Codes: LRS -- Lockheed's DIALOG
SDC -- System Development Corporation's ORBIT IV
BRS -- Bibliographic Retrieval System
COSMOS -- San Mateo Educational Resource Center's COSMOS

--- I need a positional operator between ADJ and WITH. BRS

Reviewing the positional operator commands among the three systems points out that LRS' (nW) operator would provide this function. On the other hand, LRS lacks a sentence positional operator like WITH. LRS

* * *

These comments point up the need for a review of how ERIC online searching is affected by different features of retrieval systems. The following pages analyze these differences across systems and insert additional suggestions from online searchers.

Table 3. Analysis of Existing Features in Retrieval Systems

Capability for Word Adjacency, Word Proximity, Sentence Field, and Citation Searching in Three Retrieval Systems

Retrieval System	Positional Operator	Example
Word Adjacency		
BRS	ADJ	NEEDS ADJ SENSING
LRS	(W)	NEEDS(W)SENSING
SDC	STRS	STRS (TI,AB,ST) :NEEDS#SENSING:
Word Proximity		
BRS	None	---
LRS	(nW)	IDENTIFICATION(3W)GIFTED
SDC	None	---
Sentence Search		
BRS	WITH	MUSICAL$2 WITH GIFTED
LRS	None	---
SDC	SENS	SENS (TI,AB,ST) :MUSIC: AND :GIFT:
Field Search		
BRS	SAME	IDENTIFS SAME GIFTED
LRS	(F)	IDENTIFICATION(F)GIFTED
SDC	None	---
Citation Search (anywhere in one record)		
BRS	AND	ATHLETICS AND INJURIES
LRS	(C)	ATHLETICS(C)INJURIES
SDC	AND	ATHLETICS AND INJURIES

D. Comments on Searching Features*:

All three commercial systems offer word adjacency. However, SDC's string-search is cumbersome in terms of syntax and time-consuming because of overflow messages. Stringsearching a set of 5,000 postings may require the searcher's response to over 20 overflow messages. Searchers have commented on the parentheses which must enclose DIALOG positional operators, stating that shifting and typing characters from the top keyboard row necessitates extra steps. On the other hand, BRS' ADJ requires two additional characters. COSMOS at SMERC offers the most flexibility for free text adjacency search, demanding that the searcher leave a space between terms (e.g., NEEDS SENSING), but then descriptor searching is complicated by the designation /DE or /E which must close the statement (e.g., TALENT IDENTIFICATION/E).

LRS is the only system offering word proximity, but provides no sentence search. Field searching is allowed by SDC (e.g., ALL IDENTIF: AND GIFTED/TI), but the searcher must input the statement several times closed by the field name rather than employ a single operator.

By contrast, SMERC's COSMOS allows the searcher to input terms which can be retrieved from a sentence or field in any order. For example:
IDENTIFICATION GIFTED/F,TI would retrieve a title "Identification of Creative and Gifted Children" and "The Gifted and Talented: Identification Tests."

---

*For purposes of comparison, the SMERC online retrieval system (COSMOS) will be cited to show features not available on the three commercial systems. COSMOS designers had the benefit of hindsight as it is the latest of the four systems to be designed. Please note that changes are still occuring in the design of all the systems.

The following tables indicate retrieval system features for search and print related tasks, and housekeeping functions. They point our what systems offer these capabilities, and how differently these functions are specified from system to system. A seemingly simple task, selecting a descriptor varies among tne three systems:

BRS	1_:	SABBATICAL-LEAVES
LRS	?	S SABBATICAL LEAVES
SDC	USER:	SABBATICAL LEAVES

Our studies and the work of many others cited earlier point to a clear need for some standardization of retrieval system command language. ANSI has shown some interest in this area and is about to form a Z39 subcommittee with the senior author of this project report to sit in the chair.

Table 4. Comparison of Retrieval System Commands for Searching, Printing and Housekeeping Tasks

SEARCH RELATED TASKS	BIBLIOGRAPHIC RETRIEVAL SERVICES	LOCKHEED INFORMATION SYSTEMS	SYSTEM DEVELOPMENT CORPORATION
Enter search mode	R0201 ENTER COMMAND _: ..SEARCH BRS-SEARCH MODE-ENTER QUERY 1_: GIFTED	? SELECT GIFTED (or) ? S GIFTED	(Always in search mode; to enter search disguised as a command when ORBIT asks a question, use command FIND): FIND GIFTED (or) FD GIFTED
Select search term	1_: GIFTED	? S GIFTED	USER: GIFTED
Select multi-term descriptor or identifier	1_: TALENT-IDENTIFICATION	? S TALENT IDENTIFICATION	USER: TALENT IDENTIFICATION
Select single-term descriptor or identifier	1_: IDENTIFICATION.MJ,MN.	? S IDENTIFICATION/DF	USER: IDENTIFICATION/IT
Select adjacent free text terms	1_: GIFTED ADJ CHILDREN	? S GIFTED(W)CHILDREN	USER: GIFTED AND CHILDREN PROG: SS 1 PSTG (98) USER: STRS 1 (TI,AB): GIFTED#CHILDREN:
Select adjacent free text terms with truncation	1_: GIFTED ADJ CHILD:		USER: GIFTED AND ALL CHILD: PROG: SS 1 PSTG (112) USER: STRS 1 (TI,AB): GIFTED#CHILD:
Select terms using word proximity		? S GIFTED(3W)CHILDREN	
Select terms in same sentence	1_: GIFTED WITH CHILD$3		USER: GIFTED AND ALL CHILD: PROG: SS 1 PSTG (112) SENS 1 (TI,AB) :GIFTED#CHILD:
Select terms in same field	1_: GIFTED SAME CHILDREN	? S GIFTED(F)CHILDREN	USER: (TI)GIFTED AND CHILDREN
Select terms in same field with truncation	1_: GIFTED SAME CHILD$3		USER: (TI)GIFTED AND ALL CHILD:
Positional and logical operators	1_: LOCUS ADJ CONTROL AND WOMEN	? S LOCUS(W)CONTROL AND WOMEN	
Nesting with positional and logical operators	1_: LOCUS ADJ CONTROL AND DEPRESS$ AND (WOMEN GIRL$ FEMALE$)	? S LOCUS(W)CONTROL AND DEPRESS? AND (WOM?N OR GIRL? OR FEMALE?)	
Nested boolean and set numbers	1_: LOCUS ADJ CONTROL AND 1 AND (WOMAN GIRLS FEMALE$)	? S LOCUS(W)CONTROL AND #1 AND (WOM?N OR GIRL OR FEMALE)	
Select terms with truncation using field delimeters	1_: GIFTED ADJ CHILD$/TI,ID		USER: GIFTED AND ALL CHILD:/TI,ID

SEARCH RELATED TASKS	BIBLIOGRAPHIC RETRIEVAL SERVICES	LOCKHEED INFORMATION SYSTEMS	SYSTEM DEVELOPMENT CORPORATION
Author searching	1_: SMITH ADJ (R OR RICHARD)	? S AU=SMITH, R?	USER: SMITH, R:/AU
Select year of publication	2_: ..LIMIT/1 YR GT 77	? S YR=77; S YR=78 ? C 1 AND 2	USER: 1 AND FROM 77 THRU 78
Unlimited character truncation	1_: IDENTIF$	? S IDENTIF$	USER: ALL IDENTIF:
Embedded character truncation		? S NAVA?O	USER: ALL NAVA#O
Limited character truncation	1_: TALENT$2	? S TALENT? ?	USER: ALL TALENT##
View online alphabetic display	1_: ROOT IDENTIF	? E IDENTIF	USER: NBR IDENTIF
View online thesaurus		? E (IDENTIFICATION)	
Save search permanently	_: ..SAVE PS (4-letter word of your choice) _: ..SAVE IDEN	? END/SAVE	
Save search for one day	_: ..SAVE (4-letter word of your choice) _: ..SAVE IDEN	? END/SAVE (searcher must .RELEASE the Search Save)	USER: SAVE
Execute a saved search	_: ..EXEC IDEN	? .EXECUTE (Serial number)	USER: RECALL (saved search name) RECALL IDENTIFICATION
Release permanent/temporary saved search	_: ..PURGE IDEN	? .RECALL IKLM ? .RELEASE IKLM	USER: PURGESEARCH IDENTIFICATION
Review online interaction	_: ..DISPLAY ALL (or) _: ..DISPLAY 15-25	? DISPLAY SETS (or) ? DS	USER: HISTORY (or) HIS 15-25
Limit to CIJE or RIE subfile	_: ..LIMIT/1 AN EQ ED _: ..LIMIT/1 AN EQ EJ	? LIMIT 1/ED (or) ? LIMIT 1/EJ	USER: 1 AND RIE/RS (or) 1 AND CIJE/FS

PRINT RELATED TASKS	BIBLIOGRAPHIC RETRIEVAL SERVICES	LOCKHEED INFORMATION SYSTEMS	SYSTEM DEVELOPMENT CORPORATION
Direct access	1_: ED136814.AN. RESULT 1 2_: ..PRINT 1 ALL/DOC=1	? TYPE ED136814/3	USER: ED136814/AN PROG: SS 1 PSTG (1) USER: PRT SS 1
Online prints, full format, five prints	_: ..PRINT 1 ALL/DOC=1-5	? TYPE 1/5/1-5	USER: PRT 5 SS 1 FULL
Online prints, bibliographic citation, two prints	_: ..PRINT 1 BIBL/DOC=1-2	? T 1/3/1-2	USER: PRT 1 SS 1
Online prints, specified fields, three prints	_: ..PRINT 1 AU,TI,MJ,MN/ DOC=1-3		USER: PRT 3 SS 1 AU,TI,IT
Receive online explanations of print commands or formats		? ?TYPE	USER: EXPLAIN PRINT

HOUSEKEEPING FUNCTIONS	BIBLIOGRAPHIC RETRIFVAL SERVICES	LOCKHEED INFORMATION SYSTEMS	SYSTEM DEVELOPMENT CORPORATION
Change or choose data base	_: ..CHANGE/ERIC	? BEGIN 1 (or) ? B1	USER: FILE ERIC
Change command names			USER: RENAME FIND TO F
Help!		? ?EXPLAIN	USER: HELP
Order Documents			USER: FILE ORBIT; ORDER ERIC
Available data bases	_: ..CHANGE/NEWS	? ?FILES	USER: FILES?
Unit record description		? ?FIELDS1	USER: EXPLAIN ERIC
Data base description		? ?FILE1	USER: EXPLAIN ERIC
System news	_: ..CHANGE/NEWS	? ?NEWS	USER: NEWS
Obtain amount of connect time since sign-on to retrieval system	_: ..TIME	? END	USER: TIME I

Delete all search statements	_: ..PURGEALL	? BEGIN1	USER: ERASEALL
Delete select number of search statements	_: ..PURGE 1-9 _: ..PURGE 12-24		USER: KEEP 10, 11, 25
Change length of system messages	_: ..SET DETAIL=ON (or) _: ..SET DETAIL=OFF		USER: TERMINAL MESSAGE LENGTH TUTORIAL (or) TERM ML SHORT
Disconnecting from the system	..OFF	END LOGOFF	STOP

E. Comments with Suggestions for Improvement of Vocabulary Browsing Features

1. Need to improve the ability to select from the online dictionary and thesaurus.

BRS' ROOT and SDC's NEIGHBOR commands present no more than a viewing of the inverted file. Unlike LRS' EXPAND display, the searcher cannot select, combine, or expand from the display. LRS, on the other hand, in its online dictionary (basic index), is limited to postings from all free text fields. For example, expanding ATHLETES in the online dictionary produces a display with terms from all basic index fields with no breakdown (E5).

```
  E ATHLETES
REF   INDEX-TERM                TYPE ITEMS RT
E1    ATHLETAE ------------------    1
E2    ATHLETE -------------------   82
E3    ATHLETE EVANCELISTS --------    1
E4    ATHLETE'S -----------------   10
E5   -ATHLETES -------------------  367   2
E6    ATHLETES DESIGNATED
        POSITIONS ----------------    1
E7    ATHLETIC -------------------  1029
E8    ATHLETIC ACTIVITIES --------    6
E9    ATHLETIC ACTIVITIES
        (DEL NOV74) --------------          1
E10   ATHLETIC ADMINISTRATION ----    1
```

Expanding ATHLETES/TI provides the following display. Note how the postings (E4) are not limited to the title fields but include postings from all basic index fields. This display can be misleading. The command has ignored the TI field code designator. When selecting single word descriptors from the online dictionary, the retrieval will not be limited to the controlled vocabulary field. To get such a result, the searcher must input the command: S ATHLETES/DE.

```
? E ATHLETES/TI
REF   INDEX-TERM                    TYPE  ITEMS  RT
E1    ATHLETAE ------------------           1
E2    ATHLETE -------------------           82
E3    ATHLETE'S -----------------           10
E4   -ATHLETES -------------------           367    2
E5    ATHLETIC ------------------           1029
E6    ATHLETIC ACTIVITIES
        (DEL NOV74) ------------                     1
E7    ATHLETICS -----------------           1530   34
E8    ATI -----------------------           39
E9    ATIC ----------------------           2
E10   ATIGHTUGHYUGGAAGHUSIT -----           1
E11   ATIGHTUMUN ----------------           2
E12   ATIGHTUUSIM ---------------           1
```

COSMOS offers the most versatile online dictionary and thesaurus display, allowing for term selection, combination, and expansion. But terms are amended in the display according to field. For example:

```
E ATTENDANCE

EXPANSION OF ATTENDANCE
 A*       RESUMES  REL  TYPE   PHRASE
 A1 . . . . . 1 . . . . WD . .ATTEMPTING
 A2 . . . . . 1 . . . . WD . .ATTEMTS
 A3 . . . . . 1 . . . . WD . .ATTENBOROUGH
 A4 . . . . . 1 . . . . AU . .ATTENBOROUGH, P.
 A5 . . . .1165 . . . . WD . .ATTEND
*A6 . . . .2744 . . . . WD . .ATTENDANCE
 A7 . . . . 468 . 14. . EX . .ATTENDANCE
 A8 . . . . . 1 . . . . ID . .ATTENDANCE AREAS
 A9 . . . . . 1 . . . . ID . .ATTENDANCE IMPROVEMENT PLAN
 A10  . . . . 1 . . . . ID . .ATTENDANCE INFLUENCES
 A11  . . . .15 . .5. . EX . .ATTENDANCE OFFICERS
 A12  . . . 209 . .7. . EX . .ATTENDANCE PATTERNS
 A13  . . . .73 . .7. . EX . .ATTENDANCE RECORDS
 A14  . . . .22 . .5. . EX . .ATTENDANCE SERVICES
 A15  . . . . 1 . . . . WD . .ATTENDANCERELATED
 A16  . . . . 3 . . . . WD . .ATTENDANCES

NEW SCREEN AVAILABLE
```

This resembles SDC's neighbor display which provides field locations and postings for the expanded term. However, no provision is made in SDC for

selecting or combining from the display. Consequently, the searcher must either type the variant forms he wishes, or select the truncated term and risk retrieving nonrelevant variant forms. In the display below, the neighbor display for BUS points out possibly unwanted items, e.g., BUSARIES, BUSBEE, etc. The searcher is now confronted with the decision to type in all variant forms, to truncate the term (e.g., BUS:) and select from the multi-meaning display, or truncate the term (e.g., ALL BUS:) and retrieve terms like BUSBOY, BUSYBODY, BUSINESS, or BUSING, or to limit the search to BUS/IW, the best approach to all descriptors and identifiers with the word bus in them (e.g., BUS DRIVER TRAINING, BUS DRIVERS, BUS GARAGES, BUS TRANSPORTATION).

USER:
NBR BUS

PROG:

POSTINGS	*TERM*
1	*BURZYNSKI, HELEN G./AU*
1	*BURZYNSKI, Peter R./AU*
39	*BUS/TI*
825	*BUS/BI*
615	*BUS/IW*

UP N OR DOWN N?

USER:
DOWN 17

PROG:

POSTINGS	*TERM*
311	**BUS/IW*
5	*BUS DRIVER TRAINING/IT*
5	**BUS DRIVER TRAINING/IT*
5	*BUS DRIVERS/IT*
1	**BUS DRIVERS/IT*
4	*BUS GARAGES/IT*
3	**BUS GARAGES/IT*
1	*BUS TRANSPORTATION/IT*
599	*BUS TRANSPORTATION/IT*
301	**BUS TRANSPORTATION/IT*
1	**BUS TRANSPROTATIO/IT*
1	*BUSARIES/BI*
1	*BUSBEE/TI*
1	*BUSBEE/BI*
2	*BUSBEE, CYRIL B./AU*
2	*BUSBOY/BI*
2	*BUSBY/TI*

2. Need for an online rotated descriptor display.

No system offers this capability yet. An idea for a display may resemble the following:

?ROTATE MUSIC

REF	INDEX-TERM	TYPE	ITEMS	RT
T1	MUSCULAR STRENGTH		150	6
T2	MUSEUMS		273	8
T3-	-MUSIC-		946	21
T4	MUSIC ACTIVITIES		401	12
T5	APPLIED MUSIC		49	5
T6	MUSIC APPRECIATION		222	2
T7	CHORAL MUSIC		51	6
T8	MUSIC EDUCATION		1197	11
T9	MUSIC FACILITIES		38	7
T10	ORIENTAL MUSIC		15	1
T11	MUSIC READING		80	1
T12	MUSIC TEACHERS		177	1

-MORE-

The search could then combine (e.g., ST3-T5,T6), select (e.g., ST3; ST12), or expand or rotate further from the display.

3. Need for an online rotated identifier display.

Let's envision what rotating H as an identifier would look like in such a display.

?ROTATE H

REF	INDEX-TERM	TYPE	ITEMS	RT
I1	FOUR H		1	
I2	FOUR H CAREER EXPLORATION PROGRAM		1	
I3	FOUR H CENTER		1	
I4	FOUR H CLUB		1	
I5	FOUR H CLUB AGENTS		1	1
I6	FOUR H CLUBS		65	
I7	FOUR H PROGRAM		3	
I8	FOUR H PROGRAMS		2	
I9	4 H		1	
I10	4 H PROGRAMS		4	
I11	4 H TEEN INTERN REPORT		1	
I12	H R 11258		1	
I13	H R 12257		2	
I14	H R 13266		1	

-MORE-

4. Need for a free text wrap-around display.

Taking the rotated identifier list one step further, we might envision a wrap-around free text display. Frequent phrases or single words in title, abstract, corporate source, and sponsoring agency fields would be gathered together and presented in a rotated display. In the example below, the WRAP ON-LINE SEARCHING/TI command would retrieve titles containing the phrase ONLINE SEARCHING. The display is based on occurrences of phrases in titles of ERIC citations. Someone choosing lines of this list could retrieve relevant items, check their indexing, and develop a controlled vocabulary search with the help of something like AID on Medline. At the present time, the system with word proximity searching gives a result but not such an automatic display.

WRAP ONLINE SEARCHING/TI

REF	INDEX-TERM				TYPE	ITEMS	RT
w1	diagramming	online	searching			1	
w2	education	online	searching	professional librarianship		1	
w3	education	online	searching	instructional module		1	
w4	ERIC	online	searching	procedures guidelines		1	
w5		online	searching	computer data		1	
w6		online	searching	canada datapac		1	
w7		online	searching	information retrieval		1	
w8		online	searching	libraries		1	

SELECT W5, W7; t1/2/1-2

EJ102954 IR500948
ON-LINE SEARCHING OF COMPUTER DATA BASES
PREWITT, BARBARA G.
JOURNAL OF CHEMICAL DOCUMENTATION, 14, 3, 115-7 AUG 74
DESCRIPTORS: ♦INFORMATION RETRIEVAL/ ♦DATA BASES/ ♦ON LINE SYSTEMS/ COMPUTER SCIENCE

EJ172221 IR505114
PROGRESS IN DOCUMENTATION: ON-LINE SEARCHING IN INFORMATION RETRIEVAL
BARRACLOUGH, ELIZABETH D.
JOURNAL OF DOCUMENTATION, 33, 3, 220-38 SEP 77
DESCRIPTORS: ♦COMPUTERS/ ♦INTERACTION/ ♦MAN MACHINE SYSTEMS/ ♦ON LINE SYSTEMS/ ♦STATE OF THE ART REVIEWS/ ♦SYSTEMS DEVELOPMENT/ INFORMATION RETRIEVAL

To explain how we established such links without the benefit of a lot of computer analyses: we noted, for example, frequent occurrences of two terms LEARNED HELPLESSNESS and DEPRESSION in a title which looked relevant and related.

REF		INDEX-TERM	TYPE	ITEMS	RT
W24		DEPRESSION	LEARNED HELPLESSNESS	2	
W25		DEPRESSION	OPPRESSED	2	
W26		DEPRESSION	RECOVERY HIGHER	1	
W42	INDUCED ELATION	DEPRESSION	SPEECH INITIAL	1	
W43	IRRATIONAL BELIEFS	DEPRESSION		1	
W44	LEARNED HELPLESSNESS	DEPRESSION		2	
W45	LEARNED HELPLESSNESS	DEPRESSION	CLINICAL MODEL	1	

Performing a brief search in DIALOG on the terms, we scanned all 17 indexing records for controlled vocabulary terms used in 300 more citations.

```
? SLEARNED(W)HELPLESSNESS AND DEPRESSI?
                  44 LEARNED(W)HELPLESSNESS
                 390 DEPRESSI?
          11     17 LEARNED(W)HELPLESSNESS AND DEPRESSI?
```

The list below summarizes major and minor descriptors and identifiers used in the 17 retrieved citations.

Descriptor	Major	Minor
Psychopathology	11	0
Psychological Studies	8	0
Psychological Patterns	6	2

Identifier	Major	Minor
Learned Helplessness	10	0
Depression	10	0

Would it then not be reasonable to consider devising linkages between the controlled vocabulary and free text titles with these words based on an analysis of indexing records from the present ERIC data base? From what we've learned about the index terms for citations on LEARNED HELPLESSNESS and DEPRESSION, we could amend the TT column in the wrap-around display for phrases containing these terms.

REF		INDEX-TERM		TYPE	ITEMS	TT
W24		DEPRESSION	LEARNED HELPLESSNESS		2	5
W25		DEPRESSION	OPPRESSED		1	
W26		DEPRESSION	RECOVERY HIGHER		1	
...						
W42		DEPRESSION	SPEECH INITIAL		1	
W43	IRRATIONAL BELIEFS	DEPRESSION			1	
W44	LEARNED HELPLESSNESS	DEPRESSION			2	5
W45	LEARNED HELPLESSNESS	DEPRESSION	CLINICAL MODEL		1	5

We use the information from the list of Thesaurus Terms to make up the TT display. Thus, expanding W24 or W44 or W45 results in the following linkage to ERIC's controlled vocabulary.

? EXPAND W24

REF	INDEX-TERM	TYPE	ITEMS	TT
T1	PSYCHOPATHOLOGY	DF	322	9
T2	PSYCHOLOGICAL STUDIES	DF	2024	5
T3	PSYCHOLOGICAL PATTERNS	DF	1074	21
T4	LEARNED HELPLESSNESS	IF	44	
T5	DEPRESSION	IF	201	

5. Need for Lead-In from Free Text to Controlled Vocabulary.

Such an online searching aid as the wrap-around display would assist the online searcher to gather variant forms, develop and supplement his/her strategy. The capability already exists in LRS to expand into other portions of the same list or shift to the online thesaurus. Thus, links to the controlled vocabulary could be made for phrases with a large number of postings and designated in the display's RT column. For example, phrase W6, ATTEMPTED SUICIDE DEPRESSION, is followed by a reference in the TT column. The searcher could either select the set in order to view portions of the full citation online, or expand to the reference number for suggested controlled vocabulary terms, TT, as shown below. In the EXPAND W6 computer response, the computer compared the words in W6, found that Suicide was a descriptor and had two related terms, which is also displayed.

```
? WRAP DEPRESSION
```

REF		INDEX-TERM		TYPE	ITEMS	TT
W1	ADJUSTMENTS NEW	DEPRESSION	HIGHER EDUCATION		1	
W2	AGE CLINICAL	DEPRESSION			1	
W3	AGING	DEPRESSION			1	
W4	APPLICATION CASE	DEPRESSION	TODAY'S YOUTH		1	
W5	APPROACHES INTENSITY	DEPRESSION	RURAL WOMEN		1	
W6	ATTEMPTED SUICIDE	DEPRESSION			1	3
W7	BOSTON EDUCATION	DEPRESSION			1	
W8	CHILDHOOD	DEPRESSION	EXPLANATION BEHAVIOR		1	
W9	CLINICAL	DEPRESSION	WEIGHT CHANGE		1	
W10	COMPARATIVE FREQUENCY	DEPRESSION	VARIOUS ADULT		1	
W11	CONTEMPORARY THEORIES	DEPRESSION	CRITIQUE COMPARISON		1	
W12	CONTROL BEHAVIORS	DEPRESSION			1	
W13	CONTROL RELATED	DEPRESSION			1	
W14	CONTROL TREATMENT	DEPRESSION			1	
W15	CUES DEPTH	DEPRESSION			1	
W16	DEFICITS ASSOCIATED	DEPRESSION			1	
W17		DEPRESSION	ACTIVITY EVALUATION		1	
W18		DEPRESSION	AUTONOMIC AROUSAL		1	

-MORE-

```
? EXPAND W6
```

REF	Thesaurus Display	TYPE	ITEMS	TT
T1	SUICIDE		111	2
T2	BEHAVIOR PROBLEMS		1202	15
T3	PSYCHOPATHOLOGY		322	9

6. Highlighting Search Term Lists.

How often have you performed a search, typed a few citations online and only wanted to view portions of the text in which lists occurred?

If the input search terms were highlighted in some distinctive way, the searcher could quickly scan phrases surrounding the flagged portions in order to assess relevance.

Here are our suggestions for such a feature based on a search for DEPRESSION in Women.

? S DEPRESSION? AND (WOM?N? OR FEMALE? OR GIRL?)
346 DEPRESSION?
7950 WOM?N?
8683 FEMALE?
2952 GIRL?
6 47 DEPRESSION? AND (WOM?N? OR FEMALE? OR FIRL?)

(a) Printing the full citation until the occurrences of all search terms, in this instance, WOMEN in the title and abstract, DEPRESSION in the abstract.

EJ177413
WOMEN, DIVORCE, FIND SUICIDE
HERMAN, SONYA J.
JOURNAL OF DIVORCE, 1, 2, 107-116 W77
TAPED INTERVIEWS FROM 200 WOMEN SHOWS DIVORCE AS A TRAUMATIC LIFE CRISIS FOR WOMEN THAT LEADS TO DEPRESSION.

(b) Printing fields in which the search terms occur, in the first citation, title, descriptor and identifier, and in the second title and abstract. If lists occur in both controlled vocabulary fields and abstract, printing might be suppressed in favor of controlled vocabulary rather than free text fields.

EJ176957
SEX DIFFERENCES IN THE EXPRESSION OF DEPRESSIVE RESPONSES ON THE BECK DEPRESSION INVENTORY
DESCRIPTORS: *PSYCHOPATHOLOGY/ *SEX DIFFERENCES/ *PSYCHOLOGICAL STUDIES/ *MALES/ *FEMALES/ RESEARCH METHODOLOGY/ TABLES (DATA)
IDENTIFIERS: *DEPRESSION/ *BECK DEPRESSION INVENTORY

EJ157959
EFFECTS OF DELAY, ATTACK, AND RETALIATION ON STATE DEPRESSION AND HOSTILITY
INVESTIGATES WHETHER AN INCREASE IN ANGER IN RESPONSE TO AN ATTACK SHOULD BE ACCOMAPNIED BY AN INCREASE IN DEPRESSION, IF THE OPPORTUNITY TO RETALIATE AGAINST THE PROVOKER SHOULD REDUCE BOTH DEPRESSION AND ANGER, AND MEASURES OUTWARD HOSTILITY OF MALES AND FEMALES AS WELL AS VERBAL HOSTILITY IN FEMALES. (AUTHOR/RK)

(c) Underline, Flag, or make some obvious distinction in order to contrast with the rest of the portion.

EJ108792 CG508026
LOCUS OF CONTROL, SELF-REPORTED DEPRESSION, AND PERCEIVED CAUSES OF DEPRESSION
CALHOUN, LAWRENCE G.; AND OTHERS
JOURNAL OF CONSULTING AND CLINICAL PSYCHOLOGY, 42, 5, 736 OCT 74
EXAMINES THE RELATION OF DEPRESSION TO LOCUS OF CONTROL AND TO THE PERCEIVED CAUSES OF DEPRESSION IN A NONPSYCHIATRIC POPULATION. FINDINGS SUGGEST THAT ADOLESCENT FEMALES TEND TO HOLD THEMSELVES MORE RESPONSIBLE THAN MALES FOR UNSTAISFACTORY PERSONAL SITUATION, AND THIS EXTENDS TO THE ATTRIBUTION OF CAUSES FOR UNHAPPY MOODS. (AUTHOR/PC)
DESCRIPTORS: *INDIVIDUAL CHARACTERISTICS/ *SEX DIFFERENCES/ *PSYCHOLOGICAL PATTERNS/ *EMOTIONAL RESPONSE/ *LOCUS OF CONTROL/ RESEARCH PROJECTS/ ADOLESCENTS/ SELF CONTROL/ MENTAL HEALTH

EJ117559
HOW THE SCHOOLS LIVED THROUGH THAT REAL <DEPRESSION> OF THE THIRTIES
CUTLER, MARILYN H.
AMERICAN SCHOOL BOARD JOURNAL, 162, 5, 37-42 MAY 75
DESCRIPTORS: *ECONOMICS/ *BOARDS OF EDUCATION/ *BOARD OF EDUCATION POLICY/ *WORKING <WOMEN> / *EDUCATIONAL HISTORY/ CAPITAL OUTLAY (FOR FIXED ASSETS)/ FEDERAL PROGRAMS/ TEACHER SALARIES/ TAXES/ ELEMENTARY SECONDARY EDUCATION
IDENTIFIERS: GREAT <DEPRESSION>

7. Changes to Existing Print Functions.

Searchers have expressed their satisfaction with DIALOG's print format numbers; however, some mentioned that offering the option for creating one's own format, such as BRS and SDC provide, would enhance the print function.

Online searchers have cited ORBIT and BRS as "chatty" in terms of their dialogue with the searcher, particularly in both systems, the print command for ERIC accession numbers. Rather than presenting an unbroken string of accession numbers, like LRS, both BRS and SDC number each sequentially, wasting valuable time and paper.

DIALOG Printout of 50 ERIC Accession Numbers

```
?T1/1/1-50
1/1/1-50
EJ187047  EJ187030  EJ186872  EJ186857  EJ186848  EJ186847  EJ186837
EJ186735  EJ186779  EJ185357  EJ185160  EJ184974  EJ184924  EJ184915
EJ184866  EJ184838  EJ184821  EJ184815  EJ184787  EJ183825  EJ183430
EJ183350  EJ183333  EJ183239  EJ183227  EJ182985  EJ182957  EJ182635
EJ181805  EJ181643  EJ181629  EJ181609  EJ181607  EJ181606  EJ181605
EJ181604  EJ181603  EJ181602  EJ181601  EJ181600  EJ181599  EJ181459
EJ179005  EJ179003  EJ177759  EJ177750  EJ177089  EJ175788  EJ173686
EJ173684
```

SDC Printout of 8 ERIC Accession Numbers

```
SS 3 /C?
USER:
PRINT AN 8
PROG:
-1-
AN  - EJ187730
-2-
AN  - EJ187719
-3-
AN  - ED159296
-4-
AN  - ED159279
-5-
AN  - ED159248
-6-
AN  - ED159236
-7-
AN  - ED159218
-8-
AN  - ED159207
```

171

This concern for saving paper and time has been voiced by BRS searchers when calling up the online dictionary by the ROOT command. Note in the print-out below how much space is taken up by the display.

```
1←:   ROOT FAREWELL
    FAREWELL$                                          11 DOCUMENTS
    FAREWELL                                            6 DOCUMENTS
    FAREWELLS

1←:   ROOT TERMINATION
    TERMINATION$                                      478 DOCUMENTS
    TERMINATION                                         1 DOCUMENT
    TERMINATION--A                                      1 DOCUMENT
    TERMINATION/FOLLOW-UP                              32 DOCUMENTS
    TERMINATIONS
```

Searchers of DIALOG mentioned intermixing ED and EJ documents when printing both online and offline citations. This could be rearranged according to publication date or alphabetically by author. DIALOG's printing of CIJE documents before RIE, particularly during online printing biases a searcher's strategy using the 'citation pearl growing' approach in favor of terminology from recent journal citations. The searcher employing the citation pearl growing approach capitalizes on culling terminology from citations assessed relevant from viewing portions of retrieved items during the online interaction.

Some consideration ought to be given for creating a <u>sample print format</u>. For example, after a complicated strategy, the searcher may wish to sample the output, e.g., print online or offline every 16th document from output of 400, save the strategy, present the results to the user and call him/her to assess relevant before printing the entire set of 400. This sample print format would also aid the citation pearl grower who wishes to sample indexing terminology used over a ten-year period for a particular topic.

What follows are other suggestions we received from online searchers:

8. Expanded Search Save Capabilities

Search features such as expanding and selecting from the dictionary and thesaurus would enhance the search save routine. In some instances, the searcher does not realize that he/she should have made provision for saving the strategy, thus, the saved portion would contain unwanted sets, particularly if no command is available for deleting statements such as SDC's KEEP command. However, SDC requires that the searcher enter a Savesearch mode which automatically erases all previously entered search statements, thus the searcher must determine whether or not to perform a search save before inputting his strategy.

Searchers we talked with expressed frustration concerning BRS' save search execute function which requires inputting when no other search statement numbers are outstanding.

Public-Search-saves which are currently in the planning stage at LRS ought to be given mnemotic serial numbers or names. This might also be extended to private LRS search saves, so that the searcher can assign a name to the save, or the system automatically assign a name, for example based on the first four letters of the first search term.

9. String Searching

Searchers familiar with searching ERIC on ORBIT cited free text searching as a cumbersome process, however, SDC does provide text searching of all fields of the unit record whether they are contained in the record for display or print

purposes. As long as the searcher can reduce set size to manageable proportions, he/she can meddle with the contents of any field, e.g., source availability, issue, notes, etc. Of course, this leaves some imagination to the searcher as to what lies within these fields and how he/she can utilize this information. One searcher recalled checking the bibliographic citations to a journal reference, in which only the item's title and pages were known; she performed a title search followed by a SENS on the source field for the range of pages.

10. Stop Word Lists

The stopword lists of BRS and DIALOG eliminate some potentially useful prepositions and conjunctions for stating relationships. Consider the search topic "A Child's Concept of If."

11. Search Site Profiles

Searching site profiles would provide service tailored to the needs of individual sites. The retrieved system would maintain a profile of the site on particular searching features, e.g., offline print sort format, Educational journals holdings, ERIC microfiche collections, etc. In this way, the searcher would not have to specify the desired sort format, unless different from the stored profile, CIJE citations would only be retrieved when the site had the particular journal in its collection. The nearest ERIC microfiche collections to the searching site would be noted on the offline print. This resembles SDC's RENAME command in which the searcher can change command names, e.g., FIND to F, for the duration of his/her search session. Site profiles may enable the

renaming of field names according to the site's preference rather than the retrieval system supplier's designations, e.g., naming all full descriptors, DF, or sponsoring agency, SA, rather than systems' MJ, IT, or SPO, SN labels.

12. Housekeeping functions

Besides the need for a KEEP command mentioned earlier, we have heard comments from users who would like to enter files by name father than by number in LRS, e.g., BEGIN ERIC.

Both LRS and SDC offer online assistance; some consideration ought to be given to incorporating a HELP AIDS ERIC or ?EXPLAIN AIDS 1 which would list print and non-print aids to accompany online searching. For example, the command in ERIC would cite publications like Thesaurus of ERIC Descriptors, or "User's Guide to ERIC," and their availability. It would be a cooperative effort between data base producer and retrieval system supplier to keep the AIDS display current.

13. Retrieval System Improvement

In general, online searchers of BRS expressed interest in being able to search the journal field. This is particularly helpful when using online data base searching as an aid for reference.

DIALOG (LRS) plans to reload the ERIC data base and eliminate basic index postings from the corporate source and sponsoring agency fields in response to searchers' requests. Some attention ought to be focused on developing negative paragraph qualifications as in BRS, e.g., CORNELL..SN, IN. as an alternative to deleting sponsoring agency and corporate source field postings from the basic index.

Searchers preferred LRS as a less verbose dialogue than either SDC or BRS. They voiced discontent about long response time involved in free text searching and inability to limit by year of publication.

14. For the data base

ERIC's coverage of dissertations in Education is unclear in terms of its comprehensiveness and criteria for selection. Including the volume, issue and page numbers for abstracts of dissertations in Dissertation Abstracts International in the text of both ERIC online and RIE unit records may aid the user who wishes to supplement the ERIC abstract with the DAI abstract.

Searchers have mentioned the need for a controlled vocabulary to designate field, age and grade level, publishing sponsoring agency code, reviewing descriptors on research methodologies, research, and evaluation. A means for publicizing descriptors recently integrated into the controlled vocabulary, especially through the online thesaurus, was of interest to searchers. More scope notes were called for by online searchers, particularly the integration of the notes into the online thesaurus, but with the option to print or suppress online printing of the notes when expanding the thesaurus display. One might find the texts of scope notes valuable in terms of subject-conveying material, and add these terms to the basic index. Implementing an EXPLODE display as in MEDLINE stimulated interest. Some searchers called for automatic matching of controlled vocabulary terms which have changed in meaning over time. Partitioning the file according to practioner and researcher interest, and according to foreign and American material in content, were cited as possibilities for future development.

Difficulty of retrieving and lack of coverage of Canadian material provided incentive for ONTERIS (Auster) to build their own data base accessed by ISIS software. They can treat in-depth specific items, e.g., research, curriculum guidelines, and their indexing language used accounts for English language differences in meaning, e.g., separate schools, integration, family of schools, for which the ERIC Thesaurus provides no lead-in terms.

ERIC's handling of geographic areas drew comments from a number of searchers. Use of country codes as in PREDICASTS files might provide a suitable means for retrieving items on a given region, e.g., designating the code for New England retrieves items on Massachusetts, Maine, Vermont, etc. in addition to the broader geographic area. (Note: The new GEO data element may satisfy this request.)

Entering machine-readable data file information into the ERIC data base was suggested. Unit records for these items might follow AACR, Chapter 9 guidelines. See Tessier report for ERIC/IR (ED 152 332), or Dodd's recommendations (Drexel Library Quarterly 13 (1977), 48-69).

Most of these suggestions, if implemented, would allow the online users to be more in the "driver's seat" than they are now. Margaret Park,* in her 1977 dissertation, felt that the optimal approach for an Inquiry Process Model in such retrieval systems is one which combines the judgmental capabilities of the inquirer with the management facilities of the computer. The computer mediates the inquiry process by presenting to the inquirer the decisions to be made, the choices (outcomes) which are available, and additional information which the inquirer may need to make the choice. Linda C. Smith,** in her 1979 dissertation, carried Park's work even further by looking into the ways in which Artifical

---

*M.K. Park, Design and Verification of a Computer-Mediated Inquiry System for Access to Instructional Resource Data Files (Doctoral dissertation, University of Georgia, 1977), (University Microfilm No. 77-29, 797)

**L.C. Smith, Artificial Intelligence in Information Retrieval Systems. (Doctoral dissertation, Syracuse University, 1979)

Intelligence can help Information Retrieval. Such work may appear to be a far cry from the redesign of the ERIC data base, but we think it is closer to it than the work in 1966 which originally laid out the RIE machine-readable format as part of the design for manipulation of the file by computer to produce printed indexes. The future work in the redesign of online data bases has to be built on what we have learned about the use of online systems by humans. The next section summarizes some of that work and Part V of this project report lists a bibliography on the subject.

IV. Free Text or Controlled Vocabulary Searching--A Review of Previous Work with a Possible Bearing on Redesign.

The comparative efficiency of searching free text and controlled vocabulary has been studied repeatedly using performance measures of precision, recall, and/or cost. Barker et.al. (3)* requested search queries from 48 participants and searched these queries with 5 formulations designed for the following combinations of data bases and subject fields:

1. Chemical Abstracts Condensates 1 and 2 (CAC-1, CAC-2)-titles
2. Chemical Abstracts Condensates 1 and 2-keywords
3. Chemical Abstracts Condensates 1 and 2-titles and keywords
4. Chemical-Biological Activities (CBAC) or Polymer Science and Technology (POST-J)-titles
5. Chemical-Biological Activities or Polymer Science and Technology-titles and digests

The user submitting the query assessed the output for relevance; recall and precision ratios were calculated in the following way:

$$\text{precision ratios} = \frac{\text{number of relevant items retrieved} \times 100}{\text{total number of items retrieved}} \%$$

$$\text{recall ratios} = \frac{\text{number of relevant items retrieved} \times 100}{\text{total number of relevant items retrieved by all formulations of the query}} \%$$

"It is apparent that, as might intuitively be expected, recall increases as the volume of searchable information per citation, increases. The recall ratio of titles-plus-digests formulations is greater than that of titles only formulations and the recall of titles-plus-keywords exceeds that of either titles-only or keywords-only." (p.25) Recall ratios cited in table below confirm that recall increases as more portions of the document are searched.

* Numbers in parentheses refer to Bibliography in Part V of this report.

Recall of searches in Chemical Files
(from Journal of Documentation 28, 1972, p. 30)

Sources	Titles (average of ratios)	Keywords (average of ratios)	Titles and Keywords (average of ratios)	
CAC-1	70.6	69.8	88.8	
CAC-2	57.4	78.0	91.7	
CAC-1 & 2	65.7	72.8	92.1	
CBAC	45.0	----	96.7	titles & diges
POST-J	59.4	----	97.4	
CBAC & POST-J	51.2	----	87.2	

It is interesting that keyword-only formulations did not exceed 50% precision. Title-only searches yielded higher precision ratios than keyword-searches. The authors offer little insight into characteristics of the data base which many have attributed in part to the following precision ratios, in fact, their comment on "the inadequacy of CAC-2 titles rather than superior keyword indexing for this half" (p. 26) when explaining recall scores among CAC-1, CAC-2 and CAC-1 & 2, contradicts the repeated performance of CAC-2 title search for precision.

Precision of searches in Chemical files
(from Journal of Documentation 28, 1972, p. 30)

Sources	Titles (average of ratios)	Keywords (average of ratios)	Titles and Keywords (average of ratios)	
CAC-1	49.7	45.7	41.5	
CAC-2	60.7	47.6	47.5	
CAC-1 & 2	53.6	47.5	43.8	
CBAC	60.1	----	36.5	titles & digests
POST-J	60.1	----	42.2	
CBAC & POST-J	61.9	----	38.8	

> The authors found that keywords were searched more quickly, thus more cheaply than titles or title and keyword searching. "The addition of keywords to titles increases search time by 12%, the additional digests by 20%." (p. 28)

Atherton(1) augmented MARC records with subject descriptions taken from either the table of contents and/or subject indexes of books, and tested this new data base by searching the MARC record or the augmented MARC record. "In effect we would be pitting 'controlled vocabulary' searching against 'free text' searching in these tests. This is not quite accurate to say because the present MARC record includes the 'title' of the book and the words in the title are searchable in the MARC portion of the BOOKS data base on ORBIT." (p. 49)

The experiment designed by Atherton compared the performance of searching the MARC-like portion of the BOOKS record (title, Library of Congress subject headings, classification number) with the BOOKS record (title, subject headings, classification number, main entry, contents terms, index terms). 90 queries in social sciences and humanities were submitted by eight users; output from both BOOKS and MARC searching was judged for relevance by users.

Results of an analysis of variance indicate that:

1. Searches conducted with BOOKS (i.e. the augmented MARC record) yield a significantly greater number of relevant documents than do searches conducted with MARC. In terms of recall, BOOKS retrieved at least twice as many relevant items, and sometimes three times as many as MARC.
2. There is no significant difference between the precision ratios of searches conducted with BOOKS or MARC.
3. Searches conducted with the BOOKS subject access system take significantly less time than do searches conducted with MARC subject access system regardless of whether the searches are in the humanities or social sciences.

The National Criminal Justice Reference Service (NCJRS) provides subject searching of its data base by index terms, by free text search of titles, annotations and abstracts (TANAB), and by a combination of index and TANAB

searching. A sample of 92 requests was selected at random and information specialists drew up formulations in either index terms or TANAB. It is noteworthy that although "the combinatory method has been favored among NCJRS searchers," Carrow and Nugent(11) restricted test searches to either index term or TANAB formulations. Relevance assessment of output was performed by impartial judges, i.e., three criminal justice graduate students. The results of the analysis are cited below.

Performance of searches on Criminal Justice

	Index terms	TANAB
Recall*	68.2	41.2
Precision	87.6	82.9

*Significant - p .01

> The authors report that "there was a significant effect of search method (p. 01) with index searches achieving higher recall than TANAB searches." (p.E11) No further details are given about the statistical test. It must be emphasized, though, that "a sample of data base items relevant to each request (the recall base) was assembled without using index terms." (p. E9) This appears to contradict the significantly higher recall ratio attained by index term searching.

Hersey et. al. (20,21) submitted twelve user questions to the Smithsonian Institution, Science Information Exchange data base, a file containing information on Notices of Research projects. A second phase of the project involved 39 queries searched by free text word retrieval and subject index codes retrieved. A combination of both free text and index code searching was not done. Output was assessed for relevance by experts rather than the original requestor.

Results of Two Study Phases for search performance
(from Journal of Documentation 27, 1971, p. 177-178)

12 questions	DC (free text)	SIE codes (index codes)
Recall	53	91
Precision	61	81

39 questions	DC (free text)	SIE codes (index codes)
Recall	66	95
Precision	81	95

An overall better performance is registered by index coded search formulations in the Hersey study. The results are somewhat similar to those arrived at by Carrow and Nugent, in their investigation of index term and TANAB searches. Both analyses did not employ a combination of free text and index term searching. Both report that free text searches require more computer time than index searches; in Hersey's study, the average free text search took 36 minutes in comparison to 12 minutes for index card searches. Both Hersey and Nugent and Carrow studies call for a combination system despite the fact that their work is limited to an evaluation of two methods for searching, rather than a combinatory approach.

Olive et. al(38) compared the effectiveness of index term and title searching in Nuclear Science Abstracts. Users who generated the original request were asked to assess the relevance of output. The average precision ratios were 45.6% for index term selection and 47.3% for title word selection. The authors calculated a relative recall, i.e., the proportion of the relevant documents selected by one system which are also selected by the other. Thus the recall of title word searching relative to index terms is 51%, index terms to title word searching is 58%. The authors found that the precision of title-word formulations decreases as title length increases, whereas comparative recall, i.e., ratio of the numbers of

relevant documents retrieved by the two systems, increases.

> In instances where the user wants high recall, the authors advocate "the use of free text search methods for matching against both words in titles and index terms in one complete operation." (p. 189)

Fisher and Eichesen ("Effectiveness of Combining Title Words and Index Terms in Machine Retrieval Searches." Nature 238, 1972, p. 109-110) compared differences in the citations retrieved from searching titles, index terms, and a combination of these. Queries were taken from actual requests, or constructed by the authors for the purpose of the experiment. Index term formulations retrieved an average of 92% relevant citations (precision), an average of 31% relevant documents (precision) were retrievable from titles alone.

> The authors conclude, though, that automated retrieval would be enhanced in terms of recall if a combination of title and index term searching is available.

Byrne (10) processed 50 queries selected at random from the library's COMPENDEX SDI service nine times using various combinations of subject conveying fields: title, abstracts, subject headings and free language descriptive terms. "We assumed that the number of returns obtained by the inverted search of this table constructed on all elements represented 100% of the pertinent documents in the data base." The combination of title and abstract formulations retrieved 75% of the total documents; abstracts alone retrieved 61%; titles, subject headings yielded 41% of the citations. The combination of subject headings and free language terms retrieved 30% of the citations. The remaining three combinations are title and free language terms, 27%; title, 22%, and subject headings, 21%. Byrne made no attempt to assess relevance, deciding that "the subjectivity introduced into the study through the personal bias of the people making the relevancy judgments would confuse the results rather than clarify them." (p. 225)

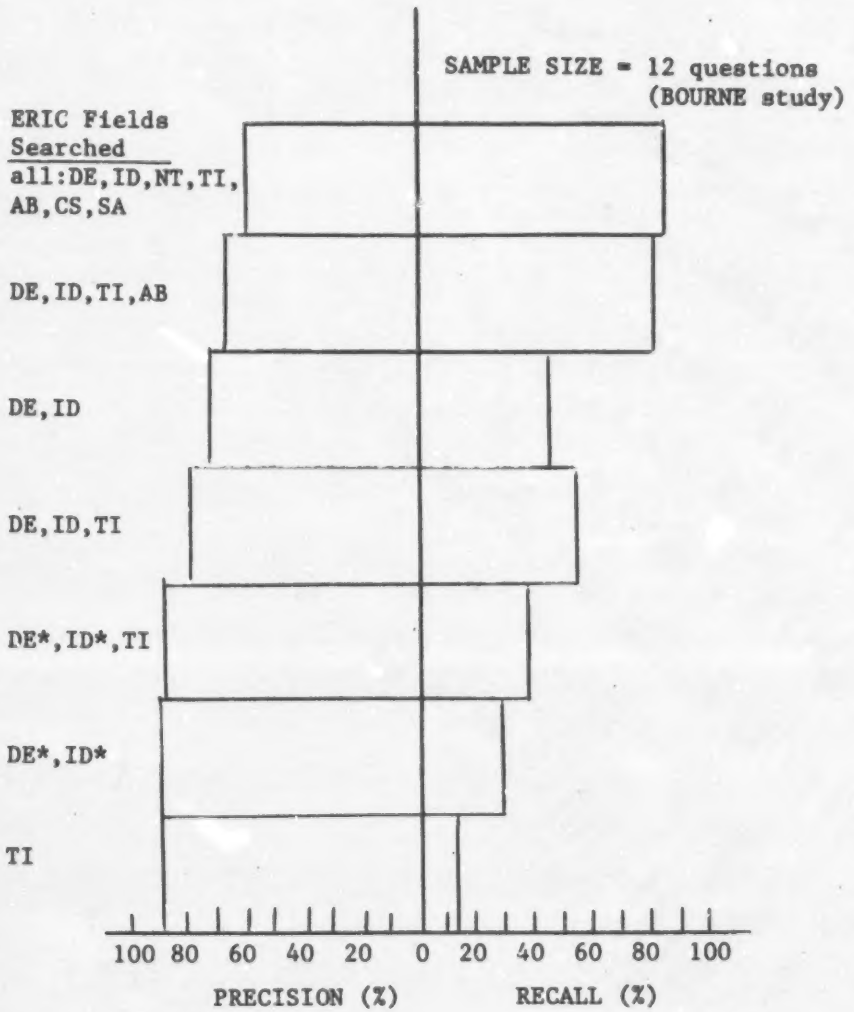
C. P. Bourne's unpublished work related to developing the ONTAP data base and its accompanying test sets, included a calculation of the percentage of total output restricted to selected titles of the ERIC record. Byrne and Bourne's studies are similar with respect to the decreasing volume of output from all subject-rich fields to title and index terms. As volume of output decreases, likewise recall decreases.

> Bourne's first figure summarizes findings in terms of recall and search output for eight queries. Note that 100% recall is not attained by any combination of fields but when searching ERIC in DIALOG, the highest possible recall is obtained by searching all subject fields. For single terms, this coincides with the default condition in which no field suffixes are used in the select command. Thus, the simplest procedure for entering searching vocabulary also gives the highest recall performance. The data presented suggests that there is little value in explicitly restricting search terms to the descriptor, identifier, title, abstract suffix combination except in cases where a subject term is likely to appear in the corporate source or sponsoring agency fields and false drops are expected. Title word searching by itself generally gives the lowest recall performance of alternatives shown.

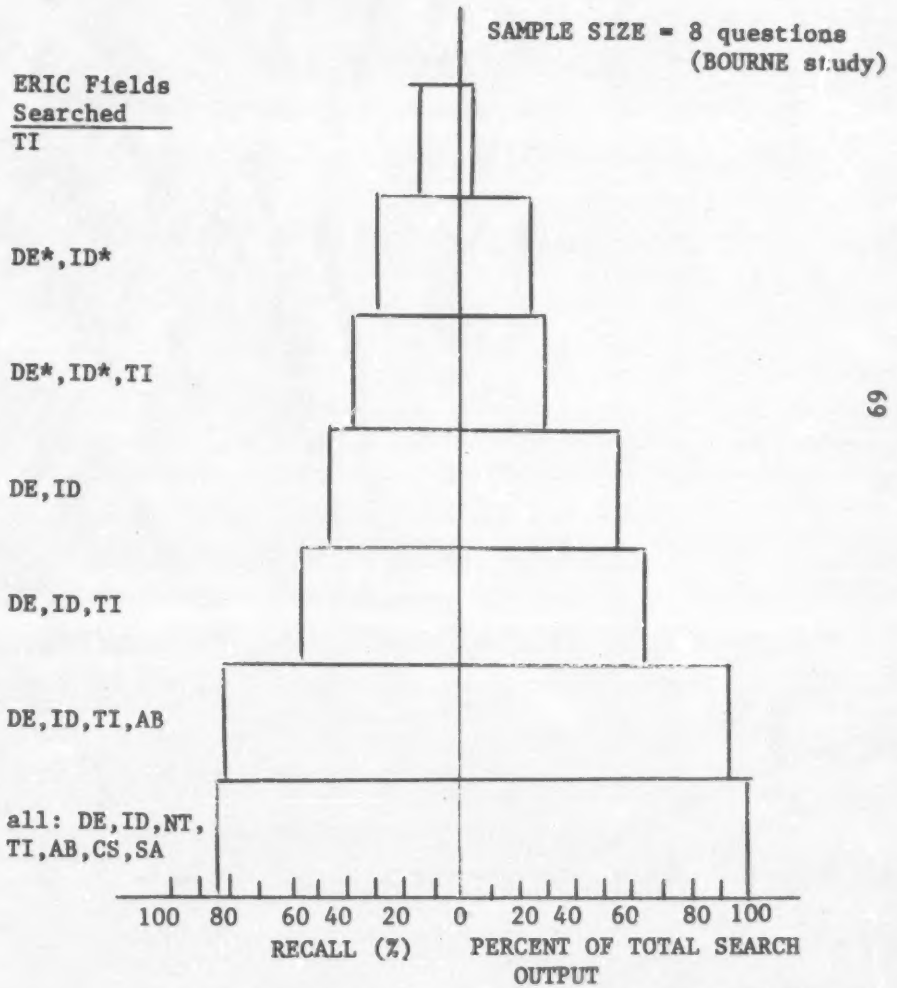
This is an interesting finding in the light of earlier work by Barker, Olive and Fisher who pitted title word searching against index term searching. Both Barker's and Olive's investigations showed that title searching yielded lower recall ratios in comparison to index term searching.

Atherton's recall ratios for MARC and BOOKS searching may be likened to index term and title searching, and full subject retrieval, respectively, in ERIC data base searching. Just as BOOKS retrieves a significantly greater number of relevant citations, Bourne's findings show a greater percentage of relevant documents when all fields were searched in comparison to index and title word searches. It is unfortunate that Hersey failed to identify what fields he considered as free text in order to compare his results with Bourne's.

BOURNE'S Tables on Searching ERIC To Check Recall, Precision, and Search Output



186



187

Bourne's figures report the recall and precision performances of DIALOG searches restricted to selected fields of the ERIC record. Whereas recall varies from approximately 13% for title searching to 85% for full subject retrieval, precision ratios range from 60% to 90%, title or index terms searches yielding up to 90% precision. Atherton found no significant difference between the precision ratios of searches conducted with BOOKS or MARC. Barker reports precision ratios which range from a high of 64.2% for title searches in POST-J to a low of 36.5% for titles and digests formulations in CRAC. Precision for index term and TANAB formulations are 87.6 and 82.9, respectively, in Carrow and Nugent's NCJRS investigation. Here it is clear that precision does not differ substantially. Lastly, Hersey's study produced precision ratios ranging from 61% to 81%, and 81% to 95% for free text and index coded formulations, in two separate experiments.

> The data in Figure 6 shows that a title search provides the highest precision of the alternative approaches tested. Searching on major index terms (DE*, ID*) achieves almost the same precision performance. And within the controlled vocabulary, there is the possibility for even greater precision by restricting the search with the IF or DF, ir IF*, DF* suffixes to full descriptors and identifiers in order to avoid false drops which occur when the index terms is used in a multi-term descriptor or identifier. By examining figures 5 and 6, one can choose a combination of data fields appropriate for a given circumstance. For example, if a high volume and high recall result are search objectives, a free text search of all subject-rich fields is in order and a range of other options are also possible. --Bourne

V. Recommendations for Future Study

The study of recall and precision for six ONTAP queries listing free text and controlled vocabulary formulations follows Bourne's findings--that recall increases, precision decreases as more subject fields are included for retrieval in the formulation, and that precision increases, recall decreases when index

term searching is compared to free text searching of all subject-rich fields. It is interesting that both our and Bourne's free text formulations failed to achieve 100% recall.

One should not overlook that the decrease in precision from title, descriptor, and identifier searches to full subject retrieval is not as great as the decrease in recall from controlled vocabulary to full subject retrieval. Often during our conversations with online searchers, false drops were associated with free text searching. The literature then suggests that additional work is needed in this area.

Our recommendations for future study are as follows:

1. Controlled time test of inputting free text formulations and controlled vocabulary formulations, particularly at various times of the day and week.
2. Replication of Bourne's tables with searching cost, time of day and week, and retrieval system as suggested variables.
3. Controlled time test of inputting descriptors one at a time in comparison to online thesaurus expand and select.
4. Controlled time test of search save execution using controlled vocabulary and free text formulations.
5. Recall and precision performances for outputs in which controlled vocabulary search saves are executed (in addition to the query formulation) and in which free text search saves are executed (in addition to the query formulation). This may answer whether the integration of a free text search save into a search profile adversely affects precision.
6. The literature on free text and controlled vocabulary retrieval effectiveness has used various groups for assessing relevance. Comparisons of relevance assessments between "impartial judges," subject experts, the original requestor, and information professionals who interviewed the requestor and performed the search are needed.
7. Comparisons of the performance in terms of recall, precision and cost for queries generated by actual library users and those orignating within an experimental environment.
8. Analysis of time involved with pre-search preparation for controlled vocabulary and free text formulations and its relationship to time spent online.
9. Do a more comprehensive bibliography on this subject than that available in Part V of this report.
10. Carry further some of work described in Linda C. Smith's dissertation concerning characteristics of document representations and possible interactive solutions for inquiry processing.

ERIC/IR
Special Project Report

ONLINE SEARCHING OF ERIC

PART V

ONLINE SEARCHING USING FREE TEXT OR CONTROLLED VOCABULARY

1970 - 1978

An Analyzed and Annotated Bibliography

by

Michael L. Rosson and Pauline Atherton

ERIC Clearinghouse on Information Resources
Syracuse University
1979

190

PART V

ONLINE SEARCHING USING FREE TEXT OR CONTROLLED VOCABULARY 1970-1978

An Analyzed and Annotated Bibliography

by
Michael L. Rosson and Pauline Atherton
ERIC Clearinghouse on Information Resources
Syracuse University

Abstract

The purpose of this annotated bibliography is to present citations that address the following topical questions about online searching using free text or controlled vocabulary search strategies:

1. What are the perceived advantages, disadvantages, and costs associated with online free text searching?

2. In which content areas are free text searches more effective than controlled vocabulary searches; and conversely, in which content areas are controlled vocabulary searches more effective than free text searches?

3. Which online bibliographic retrieval systems incorporate free text search features?

4. What are the search strategies and tactics associated with online free text and controlled vocabulary searches?

This bibliography was arbitrarily limited to citations that were: (1) published from 1970 through 1978, and (2) published in the English language.

From this analysis, we can see that 15 different data bases are mentioned by name in these references and that 24 different retrieval systems or versions of a system (e.g., ELHILL and MEDLARS, SUPARS, TEXT-PAC, STAIRS) are reported. Sixteen of these reports specifically refer to the ERIC data base.

A few reports discuss searching both by controlled vocabulary and by free text.

Some performance analysis in general (or specifically, costs, recall/precision measures, and user satisfaction) is part of just less than half of these reports.

Search strategy analysis is a part of fourteen reports.

(Contents on next page)

TABLE OF CONTENTS

Page

Abstract

I. Scope and Coverage of Bibliography 1

A. Search Strategy One 2

B. Search Strategy Two 4

C. Analysis of Bibliography's Contents 4

II. The Annotated Bibliography 10

LIST OF TABLES

Page

TABLE 1. Summary of Online Search Results 3

TABLE 2. Analytic Summary of Bibliography 6-9

I. Scope and Coverage of Bibliography

More and more, online retrieval systems allow the searcher to access the full record, choosing search terms from the title, abstract, and other fields. Besides the indexing record, which contains the descriptors (controlled vocabulary terms) chosen by the data base indexer, a searcher can now try to match the exact terms in a query and phrase search strategies in a more "natural" way. We thought it best to survey the literature of the last few years (1970-1978) to see if the "online revolution" has basically affected our approaches to bibliographic literature searching. This bibliography is a start toward such a collection of references. Volumes of ARIST and other review papers obfuscate this issue by mixing references about this topic with data base management, automated indexing and language processing, retrieval system design, etc. Here we focus on the search process itself and the variations possible because of the differences in access to the file which a list of free text terms or controlled vocabulary terms gives the searcher.

The purpose of this annotated bibliography is to present citations that address the following topical questions about online searching using free text or controlled vocabulary search strategies:

1. What are the perceived advantages, disadvantages, and costs associated with online free text searching?
2. In which content areas are free text searches more effective than controlled vocabulary searches, and conversely, in which content areas are controlled vocabulary searches more effective than free text searches?
3. Which online bibliographic retrieval systems incorporate free text search features?
4. What are the search strategies and tactics associated with online free text and controlled vocabulary searches?

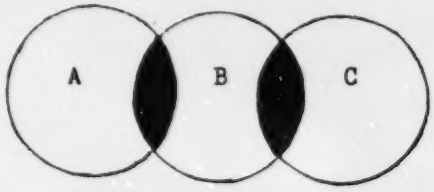
This bibliography was arbitrarily limited to citations that were: (1) published from 1970 through 1978* and (2) published in the English language.

Excluded from this bibliography are items dealing exclusively with either automatic language processing, vocabulary management, or comparisons between manual and online searches, which do not comment on differences in source of search vocabulary terms.

A. Search Strategy One

The initial search strategy was a computer search using the Educational Resources Information Center (ERIC) and the National Technical Information Service (NTIS) data bases.

The search topic was subdivided into three categories: (1) vocabulary type (concept letter A); (2) the online system (concept letter B); and (3) search strategies (concept letter C). The Venn diagram illustrated below shows the relationship between the three concepts. Any citations that were contained within the black areas of the figure were potentially desired citations.



Following the concept identification process and the formulation of the search logic, search terms were compiled. Column III of Table 1 lists the selected search terms for the two data bases. Results of the two searches are listed in Column IV of Table 1. Some forty references were found.

* Two very relevant articles were just published and are not listed here, but the reader is referred to: R.E. Hoover, "A Comparison of Three Commercial Online Vendors," Online (Jan. 1979), 12-21; S. Burrows and S. Kyle, "Searching the MEDLARS File on NLM and BRS: A Comparative Study," Bull. of the Medical Library Assn., 67 (Jan. 1979), 15-24.

TABLE 1. Summary of Online Search Results

I. Concept Letter	II. Set Number	III. Terms and Boolean Operations	IV. Number of Postings	
			ERIC	NTIS
A	1	S FREE (W) TEXT	008	004
A	2	S FULL (W) TEXT	165	007
A	3	S FULL (W) CITATION	002	000
A	4	S NATURAL (W) LANGUAGE	254	341
A	5	S KEYWORD ?	114	048
A	6	S DESCRIPTOR ?	000	000
A	7	S CONTROLLED (W) VOCABULARY	027	004
B	8	S ON-LINE	001	002
B	9	S ON (W) LINE	247	563
B	10	S ONLINE	059	035
B	11	S INTERACTIVE	990	1113
C	12	S SEARCH (W) STRATEGY	052	011
C	13	S SEARCH (W) STRATEGIES	432	085
C	14	S SEARCH (W) TACTICS	000	000
	15	C 1-7 / OR	535	399
	16	C 8-11 / OR	1219	1654
	17	C 12-14 / OR	453	094
	18	C 15 AND 16	005	005
	19	C 16 AND 17	033	011

B. Search Strategy Two

A manual search, for the years 1970 through 1978, scanned: (1) the bibliographies and references of the items retrieved in search strategy one; (2) the chapters in the Annual Review of Information Science and Technology; (3) the Proceedings of the ASIS Annual Meeting; (4) the Journal of the American Society for Information Science; (5) the Journal of Documentation; (6) Online Review; and (7) Online.

As a last step in the manual search process, all journal articles retrieved through the computer and manual searches were searched in the Social Science Citation Index.

These strategies uncovered some fifty-three relevant references which appear with annotations in this bibliography.

C. Analysis of Bibliography's Contents

An analytic summary of thirty-eight of these fifty-three citations is presented in Table 2. Excluded from Table 2 are fifteen general literature reviews (see listing below), citations of a tangential nature, and bibliographies.

Author	Document Number
Davies & Cyr	12
ERIC	15
Harris & Roberts	17
Hawkins	18
Hawkins	19
Lancaster & Fayen	29
Larson	30
Liston & Howder	32
McCarn	33
Samuelson	43
Schipma	44
Stern	47
Wanger et al.	50
Williams	51
Williams	52

From this analysis we can see that fifteen different data bases are mentioned by name in these references and that twenty-four different retrieval systems or versions of a system (e.g., ELHILL and MEDLARS, SUPARS, TEXT-PAC, STAIRS) are reported. Sixteen of these reports specifically refer to the ERIC data base.

A few reports (16) discuss searching both by controlled vocabulary (A columns in Table 2) and by free text (B columns in Table 2).

Some performance analysis (C columns in Table 2), in general, or specifically: costs, recall/precision measures, and user satisfaction, is part of just less than half of these reports.

Search strategy analysis is a part of fourteen reports (see D columns in Table 2).

We would appreciate comments and additions* Please send your comments to: Pauline Atherton, ERIC/IR, 113 Euclid Ave., Syracuse, New York, 13210.

---

* We regret the inadvertent omission of J. H. Bement, "The New Prices -- Some comparisons," Online 1(April, 1977), 9-22 and C. Cuadra's comment on Bement, Online 1 (Oct. 1977), 3,90.

TABLE 2. Analytic Summary of the Bibliography

Author's Name	Document Number	System's Name	Data Base(s) Used (See Key	Batch Processed	Online	A: Controlled Vocab. Search	A: Descriptor / Index Word Access	B: Free-Text Search	B: Title Access	B: Abstract / Digest Access	C: Performance Analysis	C: Cost	C: Precision	C: Recall	C: User Satisfaction	D: Search Strategy Analysis	D: Search Examination Analysis
ATHERTON	1*	SDC	1		X						X	X	X	X			
AUSTER & LAWTON	2		6		X												
BARKER, VEAL, & WYATT	3*	CAS	2, 3, 13								X	X	X	X			
BATES	4															X	
BATES	5															X	
BLOCK	6		4, 6, 9, 14, 15			X										X	
BOURNE et al.	7*	DIALOG	6		X											X	X
BOURNE et al.	8*	DIALOG	6		X	X		X			X	X					X
BRACKEN	9	OBIAS	6	X		X	X										
BYRNE	10*	LLL	4	X	X	X		X	X		X			X			

* See Part IV of this project report for a detailed discussion of this work.

Data Base Key

1. BOOKS
2. Chemical Abstracts Condensates
3. Chemical-Biological Abstracts
4. COMPENDEX
5. Epilepsy Abstracts
6. ERIC or ERIC/ONTAP
7. INFORM
8. INSPEC
9. MEDLARS
10. NCJRS
11. NTIS
12. Nuclear Science Abstracts
13. Polymer Science and Tech.
14. Psychological Abstracts
15. SSIE

TABLE 2. Analytic Summary of the Bibliography

						A		B			C						D	
Author's Name	Document Number	System's Name	Data Base(s) Used (See Key	Batch Processed	Online	Controlled Vocab. Search	Descriptor / Index Word Access	Free-Text Search	Title Access	Abstract / Digest Access	Performance Analysis	Cost	Precision	Recall	User Satisfaction	Search Strategy Analysis	Search Examination Analysis
CARROW & NUGENT	11*		10		X	X	X	X	X	X	X		X	X		X	
DI FONDI et al.	13	CIRC			X	X	X	X						X		X	
DOSZKOCS	14	ELHILL	9		X	X	X	X	X	X							X
FLOTT	16	RIC	6	X		X	X									X	
HERSEY et al.	20*	DATA CENTRAL	SCIENTIFIC		X	X		X			X		X	X			
HERSEY et al.	21*	DATA CENTRAL	SCIENTIFIC		X	X	X	X			X	X	X	X	X		X
HITCHINGHAM	22		2		X						X	X	X	X	X	X	
KAMINECKI et al.	23		2	X	X						X	X				X	
KATZER	24	SUPARS	14			X		X			X	X					
KIEWITT	25	PROBE	2														

* See Part IV of this project report for a detailed discussion of this work.

Data Base Key

1. BOOKS
2. Chemical Abstracts Condensates
3. Chemical-Biological Abstracts
4. COMPENDEX
5. Epilepsy Abstracts
6. ERIC or ERIC/ONTAP
7. INFORM
8. INSPEC
9. MEDLARS
10. NCURS
11. NTIS
12. Nuclear Science Abstracts
13. Polymer Science and Tech.
14. Psychological Abstracts
15. SSIE

TABLE 2. Analytic Summary of the Bibliography

Author's Name	Document Number	System's Name	Data Base(s) Used (See Key	Batch Processed	Online	A: Controlled Vocab. Search	A: Descriptor / Index Word Access	B: Free-Text Search	B: Title Access	B: Abstract / Digest Access	C: Performance Analysis	C: Cost	C: Precision	C: Recall	C: User Satisfaction	D: Search Strategy Analysis	D: Search Examination Analysis
KNIEFEL	26	QUEST	6	X		X					X	X					
KRAHMER & HORNE	27	RIC	6	X		X	X				X	X					
LANCASTER, RAPPORT, & PENRY	28	EARS	5		X						X		X	X	X		
LAWTON et al.	31	EISO	ONTERIS (6)		X						X	X			X		
MC ISAAC & OLSEN	34	WISE ONE	6		X	X	X										
MARKEY & ATHERTON	35	DIALOG	ONTAP (6)		X	X		X			X		X	X		X	X
MILLER	36	MEDLARS	9		X	X	X	X	X		X		X	X			
NEVILLE	37	GANDALF	6	X		X	X										
OLIVE et al.	38*	END	12		X	X	X	X	X		X		X	X		X	
OVERHAGE & REINTJES	39	INTREX	SCIENTIFIC			X		X									

*See Part IV of this project report for a detailed discussion of this work.

Data Base Key

1. BOOKS
2. Chemical Abstracts Condensates
3. Chemical-Biological Abstracts
4. COMPENDEX
5. Epilepsy Abstracts
6. ERIC or ERIC/ONTAP
7. INFORM
8. INSPEC
9. MEDLARS
10. NCJRS
11. NTIS
12. Nuclear Science Abstracts
13. Polymer Science and Tech.
14. Psychological Abstracts
15. SSIE

TABLE 2. Analytic Summary of the Bibliography

						A		B			C					D	
Author's Name	Document Number	System's Name	Data Base(s) Used (See Key	Batch Processed	Online	Controlled Vocab. Search	Descriptor / Index Word Access	Free-Text Search	Title Access	Abstract / Digest Access	Performance Analysis	Cost	Precision	Recall	User Satisfaction	Search Strategy Analysis	Search Examination Analysis
PALMER	40	DIALOG	6, 14		X						X	X					
RINEWALT	41	EUREKA			X			X			X	X					
ROUSE & LANNON	42	DIALOG	2,6,7,11		X	X		X								X	
SEMTURS	45	STAIRS			X	X		X									
STANDERA	46	TEXT-PAC	4	X							X	X					
TELL	48	VIRA	4, 6, 8		X			X									
TSCHUDI & MEREDITH	49	PROBE	6	X				X	X	X							
YARBOROUGH	53		6			X		X								X	

Data Base Key

1. BOOKS
2. Chemical Abstracts Condensates
3. Chemical-Biological Abstracts
4. COMPENDEX
5. Epilepsy Abstracts
6. ERIC or ERIC/ONTAP
7. INFORM
8. INSPEC
9. MEDLARS
10. NCJRS
11. NTIS
12. Nuclear Science Abstracts
13. Polymer Science and Tech.
14. Psychological Abstracts
15. SSIE

II. The Annotated Bibliography

The citations in this annotated bibliography are alphabetically arranged by the author's last name and sequentially numbered.

We would appreciate comments and additions. Please send your comments to: Pauline Atherton, ERIC/IR, 113 Euclid Avenue, Syracuse, New York 13210.

Author	Document No.	Author	Document No.
Aitchison, T.	52	Lannon, L.	42
Atherton, P.	1, 35	Larson, S.	30
Auster, E.	2	Lavelle, K.	52
Barker, F.	3, 52	Lawton, S.	2, 31
Bates, M.	4, 5	Liston, D.	32
Berul, L.	32	McCarn, D.	33
Block, M.	6	McIsaac, D.	34
Bourne, C.	7, 8	Markey	35
Bracken, P.	9	Meredith, J.	49
Byrne, J.	10, 32	Miller, W.	36
Carrow, D.	11	Neville, S.	37
Cyr, R.	12	Nugent, J.	11
Davies, H.	12	Olive, G.	38
DiFondi, N.	13	Olson, T.	34
Doszkocs, T.	14	Operating Systems	32
Elchesen, D.	52	Overhage, C.	39
ERIC	15	Palmer, C.	40
Fayen, E.	29	Penry, J.	28, 52
Fisher, H.	52	Rapport, R.	28, 52
Flott, N.	16	Reintjes, J.	39
Garfield, E.	52	Rinewalt, J.	41
Hall, A.	52	Roberts, S.	17
Hansen, I.	52	Rouse, S.	42
Harris, C.	17	Samuelson, K.	43
Hawkins, D.	18, 19	Schipma, P.	44
Hersey, D.	20, 21	Semturs, F.	45
Hitchingham, E.	22	Standera, O.	46
Horne, K.	27	Stern, B.	47
Howder, M.	32	Tell, B.	48
Ingles, R.	44	Tracy, J.	52
Kaminecki, R.	23	Tschudi, R.	49
Katzer, J.	24, 52	Veal, D.	3, 52
Kiewitt, E.	25	Wanger, J.	50
Kniefel, D.	26	Williams, M.	51, 52
Krahmer, E.	27	Wyatt, B.	3, 52
Lancaster, F.	28, 29, 52	Yarborough, J.	53

AITCHISON, T. M. (See Document #52.)

* * *

1.* ATHERTON, Pauline. Books Are For Use. Final Report of the Subject Access Project to the Council on Library Resources. (Syracuse, NY: Syracuse University, School of Information Studies, February 1978), 190 pp. ED 156131.

Findings are reported for a project designed to improve subject access to books by augmenting MARC records with subject descriptions. The BOOKS data base, consisting of humanities and social science monographs, was created and made available for online searching. The availability of suitable information in books to produce augmented subject descriptions, costs associated with creating the data base, and benefits derived from searching the data base were explored. Findings include: (1) suitable information to select subject descriptions was available in 90% of the books selected for the study; (2) cost of selection and input of subject descriptions was about $5.00 per title with an average computer storage cost of $.30 per record/month; and (3) the average cost of a search was about $14.00 for 20 minutes. A controlled test of 90 searches comparing online searching of MARC and BOOKS records showed more relevant items were retrieved using the BOOKS data base. This greater precision, together with lower costs for online searching than for MARC searching and the ability to answer some queries not possible using catalog information, were seen as the major benefits to be derived from searching BOOKS. Sample records, BOOKS user's manual, and searches conducted for the controlled tests are included in the appendices.

* * *

*See Part IV of this report for detailed discussion of this work.

2. AUSTER, Ethel and Stephen B. Lawton. "Automated Retrieval of ERIC Materials: Implications for Canadian Educators". Paper presented at the Annual Meeting of the Canadian Society of Education, Toronto, Ontario, June 6-8, 1974, 16 pp. ED 117818.

The disadvantages for Canadians using the ERIC system are: (1) fledgling Canadian Computer Information Systems are deprived of their financial support; (2) Canadian research included in ERIC cannot necessarily be retrieved by country of origin; (3) of the 200 journals indexed in Canadian Education Index (CEI), only 26 are found among the 600 in ERIC's Current Index for Journals in Education (CIJE); and (4) policies regarding the inclusion. Exclusion, and manner of indexing documents are beyond Canadian control. Solutions to some of these problems include: (a) more systematic and exclusive utilization of the ERIC system; (b) creation of a Canadian "clearinghouse", indexing all Canadian documents for ERIC; and (c) creation of a bilingual Canadian data collection and indexing system. Considering the complexity of the situation involved in creating an all-Canadian system, interim dependence on the ERIC data base searched via an existing U.S. online retrieval system is deemed necessary for one to three years.

* * *

3.* BARKER, F. H., D. C. Veal, and B. D. Wyatt. "Comparative Efficiency of Searching Titles, Abstracts, and Index Terms in a Free Text Base", Journal of Documentation, Vol. 28, No. 1, (March 1972), pp. 22-29.

(See also Document #52.)

This article describes an investigation and evaluation of the relative performance of four major United Kingdom chemical abstracts magnetic-tape data bases: Chemical Titles (CT), Chemical Abstracts Condensation (CAC), Chemical-Biological Activities (CBAC), and Polymer Science and Technology (POST).

The performance of each data base was evaluated in terms of currency and retrieval efficiency, using search profiles constructed from questions fur-

*See Part IV of this report for detailed discussion of this work.

nished by both government and industrial research groups (U.K.). The overall retrieval efficiency of the profiles were searched against data bases containing titles only, titles plus keywords, and titles plus digests. The results of the searches are presented.

* * *

4. BATES, Marcia J. Idea Tactics. (Seattle, WA: University of Washington, School of Librarianship), 26 pp.

This paper discussed techniques to assist in the generation of new concepts for conducting information searches, including manual and online systems for bibliographic and reference searches. The report considered three main areas involving search tactics: idea development, search-pattern formulation, and the need for optimum physical arrangement (locacion) of information systems. In particular, the author concentrated her presentation on the psychological and mental problems concerned with search strategies.

* * *

5. BATES, Marcia J. Information Search Tactics. (Seattle, WA: University of Washington, School of Librarianship), 47 pp. FM-30.

The primary focus of this paper is directed toward search tactics, which are a prime factor in the development of a well-conceived and efficient information search strategy. To support her contention, models of four types of search strategies are defined and briefly discussed. The 29 search tactics, which are the foundation of a sound search strategy, are listed, defined, classified, and discussed.

It was concluded that well-defined and clearly understood search tactics are practical and useful in both manual and online applications of search strategies.

BEMENT, J. (see footnote in Introduction about omission.)

BERUL, L. H. (See Document #32.)

* * *

6. BLOCK, Margaret L. Analysis of Existing Controlled Vocabularies: Women's Educational Equity Communications Network. (Princeton, NJ: Educational Testing Services, 1978).

This document concerns the online use of controlled vocabulary terms for the Women's Educational Equity Communications Network. Section 1 describes the controlled vocabulary matrix; Section 2 provides an analysis of existing controlled vocabularies for five data bases (ERIC, Psychological Abstracts, SSIE, MEDLINE, and COMPENDEX); Section 3 presents the actual matrix; and Section 4 provides an alphabetical rotated index.

* * *

7.* BOURNE, Charles P. and others. Analysis of ERIC Online File Searching Procedures and Guidelines for Searching. Final Report. (Berkeley, CA: California University, Institute of Library Research, November 1974), 156 pp. ED 101757.

In order to identify and explain the reasons for differences in average online search rates among terminal installations, the operations and performance of several facilities using the Lockheed DIALOG System for online searching of the ERIC data base were studied. Detailed examinations were made of such aspects as the DIALOG System response time as a function of the time of day or day of the week; the search commands and logic used by each of the terminal installations for their operations; the mix of complex, medium, or simple questions processed at each terminal location; and the extent and impact of the variant forms of descriptors in the file (e.g., singular and plural forms of the same term). Timing studies were performed to suggest some terminal procedures that

*See Part IV of this report for detailed discussion of this work.

could increase average online search speeds. Guidelines for searchers to consider for pre-search and terminal activities are presented at the end of the study.

* * *

8.* BOURNE, Charles P. and others. Selected Subroutines for Searching the ERIC Data Base with the Lockheed DIALOG Information Retrieval Service. (Berkeley, CA: California University, Institute of Library Research). ED 140870.

This collection of search formulations was assembled to assist in DIALOG searching of the ERIC data base. The formulations represent modular building blocks or concepts that occur frequently when searching this file. Terms were selected with high recall as the goal. Formulations are grouped into five major sections: (1) school grade level; (2) population groups; (3) types of libraries; (4) geographic areas; and (5) other topics. Posting counts are given to indicate the relative contribution of each of the terms used in the formulation. The approximate processing time and cost for execution of a search/save is also included for each formulation.

* * *

9. BRACKEN, Paula. Otis Basic Index Access System (OBIAS): A System for Retrieval of Information from the ERIC and CIJE Data Bases Utilizing a Direct Access ERIC-CIJE File. (Eugene, OR: Oregon Total Information System, 1972), 18 pp. ED 061979.

The Otis Basic Index Access System (OBIAS) for searching the ERIC data base is described. This system offers two advantages over the previous system. First, search time has been halved, reducing the cost per search to an estimated $10 on a batch basis. Second, the "Otis ERIC Descriptor Catalog", which contains all descriptors used in the ERIC records, is updated each quarter. This frequency of updating guarantees a high level of descriptor current awareness. The

*See Part IV of this report for detailed discussion of this work.

OBIAS Search System is limited to searching against descriptors, published and unpublished, in the current file. Thus, history file searches (1968 or before) must be searched with the earlier system. The OBIAS description includes: an overview of the system, system procedures, alternatives to the basic system, an example of the search card for OBIAS and statistics of run times.

* * *

10.* BYRNE, Jerry R. "Relative Effectiveness of Titles, Abstracts, and Subject Headings for Machine Retrieval from the COMPENDEX Services", Journal of the American Society for Information Science, Vol. 26, (July-August 1975), pp. 223-227. EJ 122942

(See also Document #32.)

This paper describes an investigation conducted to determine the efficiency of information retrieval by use of titles, subject headings, abstracts, free-language terms, and selected combinations of these parameters. The data employed for this study was COMPENDEX, a computer-tape version of the Engineering Index Monthly.

The results of the study indicated that a combination of titles yield the highest retrieval efficiency (75%). As a consequence of the study, the investigators concluded that a combination of the proposed retrieval technique in combination with the human-indexer method is the optimum technique for information retrieval.

* * *

11.* CARROW, Deborah and Joan Nugent. "Comparison of Free Text and Index Search Abilities in an Operating Information System", Proceedings of the ASIS Annual Meeting, Vol. 14 (1977). E8-F3.

The effectiveness of free text and index searching for the National Criminal Justice Service (NCJRS) data base was studied. A sample of 92 search requests were randomly selected, and from this sample 23 requests were drawn

*See Part IV of this report for detailed discussion of this work.

for use in a comparative analysis of free text and index searching. Index searching for the NCJRS data base achieved an average recall percentage of 68.2 and an average precision percentage of 87.6. In comparison, free text searching achieved an average recall percentage and an average precision percentage of 82.9. For the NCJRS data base the investigators recommended the use of free text searching when facets of the search contain only broad-index terms of when no index terms exist for the search facet. Furthermore, the investigators suggested that the most effective searches in the NCJRS data base are ones which combine the free text and index features.

* * *

12. DAVIES, Hopkin M. and Ralph Cyr. A Study of the Validity and Utility of the Publication Type Data Element Categorization Scheme in the Resources in Education File. (Washington, DC: ERIC Clearinghouse on Teacher Education, June 1977).

The publication type element was added to the RIE file in 1974. Since then, every addition has been assigned one of 19 codes. The Teacher Education Clearinghouse studied the utilization of the publication type codes in computer searching and the perceived validity of the code. Data sources included the 16 ERIC Clearinghouses and 16 selected intermediary service agencies, i.e., private or governmental agencies which provide information to educational personnel. Chapter headings include publication type utilization level, reasons for non-use, results of searches in which publication type codes were used as elements of search strategy, recommendations for improving code systems, the coding exercise, and conclusions/implications. Major findings entail: (1) codes used in elements in strategies in 1.5% of clearinghouse searches, 8.8% in outside agencies; (2) reasons for non-use; (3) percentage of documents assigned codes and document type descriptor; (4) recommendations; and (5) results of coding exercises.

* * *

13. DI FONDI, Nicholas M. and others. "Benefits and Costs of Free Text Searching on the FTD CIRC Reference Retrieval System", Proceedings of the ASIS Annual Meeting, Vol. 10, (1973), pp. 47-48.

An experimental study was conducted of free text and controlled vocabulary searching for the Central Information Reference and Control (CIRC) Systems. A recall experiment indicated that both searching techniques performed with approximately the same effectiveness. The free text method achieved a slightly higher total retrieval. Recall failure analysis for free text searching indicated that "12% were due to synonym problems, while 38% were caused by searching process, and 50% were attributable to use/system interface." In comparison, recall failure for controlled vocabulary indicated that "27.5% were attributable to index language and index searching process, while 49% were due to searching process, and 23.5% due to the user/system interaction." Using the data analysis as the basis for a decision, the system was totally converted to a free text indexing system. Results of this decision indicate that: (1) document input costs have been reduced five-fold; and (2) users' acceptance of the free text search strategy has been high.

* * *

14. DOSZKOCS, Tamas E. "An Associative Interactive Dictionary (AID) for Online Bibliographic Searching", Proceedings of the ASIS Annual Meeting, Vol. 15, (1978), pp. 105-109.

This paper describes an experimental prototype AID system that has been interfaced to the ELHILL retrieval software for use by the National Library of Medicine data base, including MEDLINE. The AID system uses a highly compressed "hash" table of inverted search keys, enabling the retrieval software to automatically identify conceptual text association. The online system is interactive using an association measure derived from the difference between observed and ex-

pected term (string) frequencies in retrieval sets of text segments.

The prototype AID system has the potential for automatically identifying associative linkage between chemical substructives and biological substructure in chemical-biological retrieval systems. It can be used on indexing records or abstracts of the retrieved set.

* * *

ELCHESEN, Dennis R. (See Document #52.)

* * *

15. ERIC Literature Search on Needs Assessment. (Chelmsford, MA: Merimack Education Center, 1974). ED 100382.

A search of ERIC literature was undertaken in order to assess research on educational needs. A set of ERIC descriptors was selected, and criteria for evaluation of each abstract established. Approximately 1800 abstracts were retrieved; 150 survived the first screening, and 75 the second. Priorities were then assigned to decide which documents should be retrieved. A list of the priority documents is appended with a matrix of their contents and value.

* * *

FISHER, H. Leonard. (See Document #52.)

* * *

16. FLOTT, Nancy. A Manual for Computerized Retrieval. Learning Activity Packet. (Topeka, KS: Kansas State Department of Education, 1975), 63 pp. ED 116650.

This manual deals with the mechanics of information retrieval of the ERIC files, using the RIC program of North Dakota. The material covered includes formulating a basic search strategy, identifying major concepts, and the unique

properties of the ERIC system. Each discussion contains illustrations and examples as well as a summary of important points. The appendix includes 19 worksheets, 15 answer sheets, a glossary, and a one-page bibliography.

* * *

GARFIELD, Eugene. (See Document #52.)

* * *

HALL, Angela M. (See Document #52.)

* * *

HANSEN, Inge Berg. (See Document #52.)

* * *

17. HARRIS, Curt L. and Suzanne H. Roberts. "The Search for Tomorrow: Low Cost, Full Text Search with Minimum Front End 'Investment'", Proceedings of the ASIS Annual Meeting, Vol. 15, (1978), pp. 368-372.

The paper postulates that the use of general purpose computers, consisting of hardware and software developed to perform accounting tasks, or to solve mathematical equations, are not the proper machines for the ultimate solution of the growing need for information retrieval. It is proposed that the development of a special purpose computer hardware system, Associate Processor (AP), which is implemented with highly modular software, and a minicomputer controller will result in an information retrieval system that will allow for low cost rapid searches of large bodies of texts. This development should eliminate time-consuming data base updates and provide full text searches to satisfy future information retrieval needs.

* * *

18. HAWKINS, Donald. "Online Information Retrieval Bibliography, 1975-76", Online Review, Vol. 1, No. 1, (March 1977, Supplement), pp. S1-1-S1-55.

Presented in this bibliography of 615 citations were seven topical areas dealing with online bibliographic and numeric information systems published from 1975-76. Within each section, citations were listed alphabetically by author. A permuted title and author index was provided after the bibliography.

* * *

19. HAWKINS, Donald. "Online Information Retrieval Bibliography", Online Review, Vol. 2, No. 1, (March 1978), pp. 63-106.

This was an update of a bibliography published in 1977 and contains 436 references, which were divided into the same seven major sections as in the original document. Within each section, citations were alphabetically listed by author. An author and a permuted title index was provided.

* * *

20.* HERSEY, David F. and others. "Comparison of Online Retrieval Using Free Text Words and Scientist Indexing", Proceedings of the ASIS Annual Meeting, Vol. 7, (1970), pp. 265-267.

Thirty-nine queries were utilized in four scientific content areas (pesticides, reproduction, solid state physics, and physical oceanography), to compare the effectiveness of free text and subject index searching. The results of the study indicated that the recall levels by subject index searching were 30% higher than for free text searching; and the precision levels by subject index searching were 14% higher than free text searching. The investigators concluded that: (1) free text searching should be used when detailed descriptions are needed; (2) index searching should be used when searching for broad concepts; and (3) a combined free text and index search system should be developed.

* * *

*See Part IV of this report for detailed discussion of this work.

21.* HERSEY, David F. and others. "Free Text Word Retrieval: Performance and Costs", Journal of Documentation, Vol. 27, No. 3, (September 1971), pp. 167-183.

Reported are the results of a study to determine the comparative performance for online information retrieval by (1) free text words; and (2) subject indexing codes within the content areas of reproduction, solid state physics, pesticides, and physical oceanography. The results of the study indicated that the efficiency of information retrieval was 15% to 20% higher using subject indexing codes than for free text word use.

* * *

22. HITCHINGHAM, Eileen E. Online Data Bases. II - Search Performance Evaluation and Online Data Bases. (Washington, DC: Council on Library Resources, 1975), 30 pp. ED 139388.

In an attempt to measure online searching performance, test searches of one question conducted by three bibliographic searchers are analyzed to determine the effects of search strategy on results. The study reviews retrieved citations for relevance, calculating precision values and estimated recall for each search. Performance analysis reveals that the "best" search strategy can be determined using different criteria -- low cost, better recall, or high recall. The report summarizes findings of previous studies measuring: (1) cost, terminal time, and user satisfaction; (2) precision; and (3) recall. It suggests the performance measure of unit cost per relevant citation determined by the cost of terminal time plus the cost of offline prints divided by the number of relevant citations. Suggested formulas for determining recall values in high recall need searches are defined for institutions with one searcher and institutions with several searchers.

* * *

*See Part IV of this report for detailed discussion of this work.

INGLES, R. (See Document #44.)

* * *

23. KAMINECKI, Ronald M. and others. "Searching CA Condensates, Online, and Batch". Paper presented at the Annual Meeting of the American Chemical Society, Chicago, Illinois, August 26, 1975, 13 pp. ED 121303.

Batch mode processing is compared, using cost effectiveness, with online processing for computer aided searching of Chemical Abstracts. Consideration for time, need, coverage, and adaptability are found to be the criteria by which a searcher selects a method, and sometimes both methods are used. There is a trade off between batch mode's slower access and lower hardware requirements and online's speed but greater expense. Both methodologies involve programs which are constantly changing, and therefore, require information personnel who can be retained easily.

* * *

24. KATZER, Jeffrey. "The Cost-Performance of an Online, Free Text Bibliographic Retrieval System", Information Storage and Retrieval, Vol. 9, (1973), pp. 321-329.

(See also Document #52.)

This study compared the cost effectiveness and performance of the experimental free text system SUPARS I and II. The data base (DDB) contained bibliographic records, consisting of citations and 200-300 word abstracts for 2 to 3 years of Psychological Abstracts. For SUPARS I, the estimated average computer cost for one search query, using only the OR and AND Boolean operators was $.59. Using the OR, AND, and NOT Boolean operators, and suffix truncation, the estimate average cost was $1.10. In turning to SUPARS II, the search of the document data base (DDB) cost 14% less than SUPARS I. Katzer attributed this to a change in the hardware configuration, from an IBM 360/50 to an IBM 370/155 computer.

In SUPARS II, an online thesaurus, the vocabulary data base (VDB), was added to the system. The investigator pointed out that "the use of the VDB decreases the need to search the DDB. Thus, the user can trade off expensive searches to the DDB for searches to the VDB (which are likely to remain relatively inexpensive regardless of its growth)."

* * *

25. KIEWITT, Eva L. "PROBE": Computer Search of the ERIC Tapes. (Bloomington, IN: Indiana University, 1973). ED 075050.

PROBE, a search and retrieval program, was developed in 1970 at Indiana University to search the computer tape bases for the two ERIC publications, Research in Education and Current Index to Journals in Education. The ERIC tapes are received, converted, and combined at the University. This paper reviews the early experimental stages of the program and documents the project growth in staff and number of searches conducted in response to user requests.

* * *

26. KNIEFEL, David R. "Quest: A Conversational Access to Computerized Searches of the ERIC System". Paper presented at the American Educational Research Meeting, New Orleans, Louisiana, February 25 - March 1, 1973, 18 pp. ED 074758.

Quest is an interactive conversational programming system (CPS) that was developed to serve as a conversational interface between a searcher of the ERIC files and the North Carolina Science and Technology Research Center's Inverted File Search Program (STRC-IVS). This paper describes Quest, its costs, operational procedures, and problems. The Quest Program has not only proven to be extremely successful cost-wise, but has also become a means of introducing non-computer oriented users to automated information retrieval techniques. The development of Quest has greatly increased the utilization of the ERIC files.

* * *

27. KRAHMER, Edward and Kent Horne. An Alternative to Query: Batch-Searching of the ERIC Information Collections. (Grand Forks, ND: Resource Information Center, 1972), 55 pp. ED 063519.

A manual describing the RIC Computer Search Program for retrieval of information from ERIC, CIJE, and other collections is presented. The manual places emphasis on how to utilize the RIC Program to perform searches. Consideration of some technical aspects is given in the appendices. The approach taken by RIC, utilization of the ERIC descriptor postings, provides one alternative to the Computer Search Program, Query.

* * *

28. LANCASTER, F. W., R. L. Rapport, and J. K. Penry. "Evaluating the Effectiveness of an Online Natural Language Retrieval System", Information Storage and Retrieval, Vol. 8, No. 5, (October 1972), pp. 223-245.

(See also Document #52.)

An evaluation of the Epilepsy Abstracts Retrieval System (EARS) was performed. Searches were conducted on the online system in terms of recall, precision, and general user satisfaction. A detailed analysis of factors affecting the success and failure of the online searches was carried out.

* * *

29. LANCASTER, F. W. and E. G. Fayen. Information Retrieval Online. (Los Angeles, CA: Melville Publishing Company, 1973).

In Chapter Three (pp. 37-41) the concept of online natural language retrieval of documents is presented. The advantages and disadvantages of online searching through controlled and/or natural language vocabulary are presented in Chapter Eleven. Descriptions of several online natural language systems including CIRCOL, DIALOG, EARS, ORBIT, SMART, and SUPARS are developed in various chapters of the book.

* * *

30. LARSON, Signe. "Online Systems for Legal Research", Online, Vol. 1, No. 3, (July 1977), pp. 11-14.

Three online legal data bases were presented. A brief analysis of the advantages of free text searching legal data bases is also included.

* * *

LAVELLE, Katherine H. (See Document #52.)

* * *

31. LAWTON, Stephen B. and others. A Systems Evaluation of the Educational Information System for Ontario. (Toronto: Ontario Institute for Studies in Education, 1978). ON 01154.

The purpose of this paper was to describe a study for the evaluation of online bibliographic retrieval services provided by the Educational Information System for Ontario (EISO) through the evaluation of the relationships between search procedures, the production of useful information, and the extent of user satisfaction.

Questionnaires were prepared and distributed to 371 clients of EISO to determine, in part, retrieval system used, turnaround times, connect time, number of citations printed, methods of planning for search strategy, costs of search, and relevant information produced by the search. In addition, the questionnaire probed client satisfaction with the information system by seeking an assessment of clients' attitudes about materials and directions, convenience and helpfulness, quality of technology, etc.

Results of the survey indicated that the clients' assessment of the information provided is independent of the costs of the search. But the total amount of the costs raises the user's expectations from the service provided.

32. LISTON, David M. Jr. and Murray L. Howder. "Subject Analysis", Annual Review of Information Science and Technology, Vol. 12, edited by Martha E. Williams. (White Plains, NY: Knowledge Ind. Pub., 1977), pp. 81-118.

Three citations on free text indexing/searching are included in this chapter on subject analysis. The citations are presented below:

Berul, L. H. "Natural Language Text Processing", Changing Patterns in Information Retrieval, Proceedings of the National Information Retrieval Colloquium, 1973; (Philadelphia, PA, Washington, DC, American Society for Information Science, 1974), pp. 80-85.

Byrne, J. R. "Relative Effectiveness of Titles, Abstracts, and Subject Headings for Machine Retrieval from the COMPENDEX Services", Journal of the American Society for Information Science, Vol. 26, (July-August 1975), pp. 223-227. EJ 122942. (See Document #10.)

Operating Systems, Inc. A Comparative Evaluation of Structured and Free Text Searching of the NHTSA Data Base. (Washington, DC: National Highway Traffic Safety Administration, 1975). (Report No. P-47-143).

* * *

33. McCARN, Davis B. "Online Systems-Techniques and Services", Annual Review of Information Science and Technology, Vol. 13, edited by Martha E. Williams. (White Plains, NY: Knowledge Industry Publications, Inc., 1978), pp. 85-124.

An analysis of full text bibliographic systems is presented in two sections of the chapter. The first section of interest (p. 95) presents LEXIS and WESTLAW, two online full text legal systems. Evaluation of controlled vocabulary and free text research, published during 1976 and 1977, are presented on pp. 104-106. A bibliography of approximately 200 citations, covering the years 1976 and 1977, are included at the end of the chapter.

* * *

34. McISAAC, Donald N. and Thomas Olson. "Retrieval of ERIC Files. An Online Approach". Paper presented at the Association for Data Systems Annual Convention, New Orleans, Louisiana, April 16-19, 1973, 10 pp. ED 087454.

A description is provided for WISE ONE, an information retrieval program designed to provide fast, efficient access to computer-based information files. The author focuses upon WISE ONE's application to the ERIC data base. WISE ONE was designed specifically to meet the needs of researchers using ERIC, but its logic is sufficiently general to accommodate other computer-based library systems. The author first reviews ERIC's history, its products, and its thesaurus. Following this, the binary tree and index sequential file approaches to data structuring are discussed, and their disadvantages brought to light. Next, the hash coding method of data base entry is described; the conclusion is reached that it provides the best approach to the data base. Finally the use of Boolean operators is discussed and means of updating files are considered.

* * *

35. MARKEY, Karen and Pauline Atherton. ONTAP: Online Training and Practice Manual for ERIC Data Base Searchers. (Syracuse, NY: ERIC Clearinghouse on Information Resources, Syracuse University, June 1978), 198 pp.

This manual provides the experienced online searcher with an opportunity to improve his/her skills in relation to the total search process. The first section outlines: (1) the definition of the client's search topic; (2) the identification of relevant data bases and (3) search terms; (4) approaches to search strategy; (5) knowledge of search system's features; and (6) the evaluation of preliminary and final results. Two to three search strategies for 15 topics are detailed in the second section of the manual and are specifically designed for either high recall or high precision. The search formulations are constructed either in free text, or a combination of free text and controlled vocabulary. Included in the appendix are 13 search saves for population and tool concepts.

* * *

36. MILLER, William L. "The Efficiency of MEDLARS Titles for Retrieval", Journal of the American Society for Information Science, Vol. 22, (September-October 1971), pp. 318-321.

Described is a title-search, developed for the MEDLARS System, a technique that allows the user to fix the number of references to be retrieved prior to the start of a search. The proposed title-search technique treats each title as a character sequence rather than a word sequence. The developed program allows for entry to a title through construction of a Key Letter In Context Index (KLIC).

The retrieval performance of the proposed title search tactic was compared with the standard MEDLARS Boolean and the Scored searches of index terms. Analysis of test results indicated that the proposed program yields slightly better retrieval results than either of the MEDLARS or scored techniques.

* * *

37. NEVILLE, Sandra H. Searching the ERIC Data Base at UNM. (Albuquerque, NM: New Mexico University, Zimmerman Library, December 1974), 19 pp. ED 100339.

The University of New Mexico (UNM) general library supports the locally-managed ERIC Search Service. During the 1973-74 fiscal year, more than 3,000 searches were processed by the UNM Division of Computer and Information Science using a specially developed retrieval program: GANDALF (General Alpha-Numeric Direct Access Library Facility). This document is intended as a guide to users of the ERIC/GANDALF System. Included is an overview of the system, instructions for formulating a search, specific procedures for submitting the search, and a guide to interpreting the search results.

* * *

38* OLIVE, G. and others. "Studies to Compare Retrieval Using Titles with That Using Index Terms", Journal of Documentation, Vol. 29, No. 2, (June 1973), pp. 169-191.

This paper describes a study conducted for the purpose of comparing the retrieval efficiency using titles with the retrieval efficiency using index terms. The experiment was carried out using a Selective Dissemination of Information Service (SDI) based on computer scanning of Nuclear Science Abstracts (NSA) tapes.

To compare the effectiveness of the index terms (Euratom) with free text titles, an experiment was constructed whereby 60 users of the SDI-NSA system were selected, at random, to participate in the experiment without being aware of the type of method used to produce the retrieval service.

Results of the study indicated that the SDI can operate with comparable efficiency when using index terms on free text. In particular, the same degree of precision was attained for titles having 100 characters or more as for index terms.

* * *

OPERATING Systems, Inc. (See Document #32.)

* * *

39. OVERHAGE, Carl F. J. and J. Francis Reintjes. "Project Intrex: A General Review", Information Storage and Retrieval, Vol. 10, No. 5/6, (May/June 1974), pp. 157-188.

This paper reviews the implementation, the results, and conclusions derived from the search activities of Project Intrex, an experimental information storage and retrieval system developed by researchers at the Massachusetts Institute of Technology.

An online real-time interactive computer-assisted information retrieval

*See Part IV of this report for detailed discussion of this work.

system was designed and implemented for experimental studies. A computer-stored catalog of scientific articles, related to five fields of engineering, was assembled and stored in the computer's memory. This information base was supported by 20,000 free text articles sorted in the system's microfiche files, allowing the user to have immediate access to both video display and hard copy.

A series of experiments were conducted to evaluate factors affecting the system's retrieval effectiveness and to determine the value of the availability of guaranteed access to free text displays.

In part, results of the studies indicated the free text and controlled vocabulary produced essentially the same degree of retrieval effectiveness, leading the researcher to recommend the use of free text vocabulary as the primary indexing technique in similar information retrieval systems. Also, it was found that the typical user spent approximately 83% of his/her time scanning free text displays.

* * *

(40.) PALMER, Crescentia. A Comparison of Costs of Searching the Machine-Readable Data Bases ERIC and Psychological Abstracts in an Annual Subscription Rate System Against Cost Estimated for the Same Searches Done in the Lockheed DIALOG System and PASAT for Psychological Abstracts. (Albany, NY: New York State Education Department, May 1976), 28 pp. ED 127907.

A comparison was made of costs for computer searching of the Psychological Abstracts and the ERIC data bases. Charges for peripheral equipment, which are included in the contract fee, were prorated according to the ratio of use of each data base against the total use for all data bases. Estimates were obtained for comparable peripheral equipment, adjusted for the percentage of time used, and added to search costs to complete the comparison.

* * *

PENRY, J. K. (See Documents #28, #52.)

* * *

RAPPORT, R. L. (See Documents #28, 52.)

* * *

41. RINEWALT, J. R. "Feature Evaluation of a Full Text Information Retrieval System", Online Review, Vol. 1, No. 1, (March, 1977), pp. 43-52.

The full text information system, EUREKA, was evaluated through a series of feature evaluation experiments conducted in 1975 and 1976. Cost and user performance are the criteria used in feature evaluation.

* * *

42. ROUSE, Sandra H. and Laurence W. Lannon. "Some Differences Between Three Online Systems: Impact on Search Results", Online Review, Vol. 1, No. 2, (June 1977), pp. 117-132.

A comparison of three major bibliographic search systems: BRS (Bibliographic Retrieval Services), LIS (Lockheed Information Systems), and SDC (System Development Corporation), was made in terms of their different file structures, data fields, and search features. Through analyses of postings of CA Condensates, ERIC, INFORM, and NTIS, the authors concluded that differences in dictionary files, stop words, term proximity, and subject searching account for the major differences in the strategy and results obtained from the three search systems.

* * *

(43.) SAMUELSON, Kjell. "free TEXT INDEXing to meet COMPUTers halfway", IEEE Transactions on Engineering Writing and Speech, Vol. EWS-13, No. 2, (September 1970), pp. 46-48.

Free text searching is gradually being combined with parallel processing computers and optical character reading (OCR), and the result will be a growing need for more efficient source indexing and text compression of abstracts.

A methodology is described which could be employed to improve the free text searching that should result in more efficient information retrieval. In part, the technique is based on grammarless text handling and the use of capital letters for only primary message strings (e.g., SYSTEMs ANALYsis TECHNiques).

* * *

(44.) SCHIPMA, Peter B. "Generation and Uses of Machine-Readable Data Bases", Annual Review of Information Science and Technology, Vol. 10, edited by Carlos A. Cuadra and Ann W. Luke. (Washington, DC: American Society for Information Science, 1975), pp. 237-271.

A brief analysis of free text software is developed on pp. 247-249 and a more detailed presentation of free text system research is presented on pp. 256-258. Accompanying this chapter is an 184-citation reference list, which is alphabetically arranged and sequentially numbered. The following reference specifically deals with free text effectiveness:

Ingles, R. "Text-Pac System Enhancements", American Society for Information Science, Western Canada Chapter. (Alberta, Canada: 1973), pp. 32-43.

* * *

(45.) SEEGERS, Friedrich. "STAIRS/TLS: A System for Free Text and Descriptor Searching", Proceedings of the ASIS Annual Meeting, Vol. 15, (1978), pp. 295-298.

This article describes how a retrieval system like the Storage and Information Retrieval System (STAIRS/TLS), which provides free text searching in a multi-user and multi-data base environment, can be improved.

The techniques proposed for system improvement included functional improvements such as: (1) the development and thesaurus menus for selection by the user and (2) the development of multflingual languages for international communication networks. The article discussed the need to provide: (a) descriptor-oriented retrieval functions, combined with thesaurus usage at retrieval time; (b) the development of specific functions for retrieving multilingual data bases; and (3) development of an algorithm to account for grammatical rules affecting a stored text.

* * *

(46.) STANDERA, Oldrich. COMPENDEX/TEXT-PAC: Retrospective Search. (Alberta, Canada: Calgary University, 1971). ED 051862.

The TEXT-PAC system is capable of generating indexes and bulletins to provide a current information service without the selectivity feature. The manual search involves searching computer-prepared indexes from a machine-readable data base produced largely by human abstracting and indexing. Retrospective searching in the TEXT-PAC System provides computer matching of a machine-readable data base against questions prepared manually and translated into the system language. The entire record is scanned for the occurrence of the question words and the "hits" from this matching are obtained in the form of a printout.

* * *

(47.) STERN, Barrie T. "Evaluation and Design of Bibliographic Data Bases", Annual Review of Information Science and Technology, Vol. 12, edited by Martha E. Williams. (White Plains, NY: Knowledge Industry Publications, 1977), pp. 3-33.

Free text research using precision and recall analysis is discussed on pp. 10-13 of the chapter.

* * *

48. TELL, B. "SDI Obtained from Numerous Tape Services". Paper presented at the IATUL International Association of Technical and University Libraries Fifth Triennial Conference, Copenhagen, May 1973, 13 pp. ED 089657.

The information service performed by the Royal Institute of Technology handles requests from both university users and from industry. Fourteen data bases are currently used for SDI purposes. The system is of a general nature which permits inclusion of various data bases of completely different tape formats. Any combination of elements of a bibliographic record can be searched. As the profile maintenance program becomes an essential element when the number of profiles is building up, an online facility for updating has been installed. Special emphasis is placed on the principles for free text searching and on the problems of ordering the printout in such a way that it gives user satisfaction. At present a statistical approach is in operation. Since many data bases have keywords or other subject indicators, a combination of free text search of titles and keywords is often used. There is no significant deviation of the user's reaction when given printout from free text searches or from keyword searches.

* * *

TRACY, Jennifer M. (See Document #52.)

* * *

49. TSCHUDI, Ronald R. and Joseph C. Meredith. The "PROBE" Retrieval Program: A Description. (Park Forest South, IL: Governors State University, 1972), 11 pp. ED 059596.

PROBE is a search and retrieval program designed for use with large tape files of bibliographic data such as the ERIC System's Research in Education and Current Index to Journals in Education data bases. The program is characterized by a broad range of capabilities and options which are described in this report.

The general capabilities of PROBE are briefly described, and further sections discuss PROBE search procedures and specifications. Two separate levels of search parameter specification -- Spec and Query -- are described in more detail. A brief section on the application of the PROBE program concludes the report.

* * *

VEAL, Douglas C. (See Documents #3, #52.)

* * *

50. WANGER, Judith and others. Impact of Online Retrieval Services: A Survey of Users, 1974-75. (Santa Monica, CA: System Development Corp., 1976), 292 pp.

Over 1,250 users of ten major bibliographic information retrieval services were surveyed to assess the impact of such services on the library and information science community and its user community. The major topics considered in analyzing responses were: (1) respondent backgrounds; (2) methods of acquiring and promoting online services; (3) selection and training of staffs; (4) levels of use; (5) selection, access, and use of online systems; (6) selection and use of data bases; (7) costs; (8) major problems; and (9) major areas of impact. It was found that the services are being used primarily by libraries and information service centers in commercial organizations, universities, and federal government organizations. Furthermore, the use of search services is generally considered cost-effective.

* * *

51. WILLIAMS, Martha E. "Networks for Online Data Base Access", Journal of the American Society for Information Science, Vol. 28, No. 5, (September 1977), pp. 247-253.

This paper discusses network sharing of resources for online data base searching. Uses of online searching, equipment, personnel, search strategy development requirements, costs, and advantages of such services are also considered.

* * *

52. LLIAMS, Martha E. "Use of Machine-Readable Data Bases", Annual Review of Information Science and Technology, Vol. 9, edited by Carlos A. Cuadra and Ann W. Luke. (Washington, DC: American Society for Information Science, 1974), pp. 221-277.

Contained in the chapter is a two-page summary (pp. 233-234) of investigations pertaining to the use of controlled and/or natural language vocabulary for the searching of online data bases. Accompanying this chapter is a 227-citation reference list, alphabetically author-arranged and sequentially numbered. The following references deal specifically with keyword and/or free text effectiveness.

Aitchison, T. M., Angela M. Hall, Katherine H. Lavelle, and Jennifer M. Tracy. Comparative Evaluation of Index Languages. Part II: Results. (London, England: INSPEC, July 1970), 64 pp., Report No. R70/2.

Aitchison, T. M. and Jennifer M. Tracy. Comparative Evaluation of Index Languages. Part I: Design. (London, England: INSPEC, July 1969), 1 v. (unpaged), Report No. R70/1.

Barker, Frances H., Douglas C. Veal, and Barry K. Wyatt. "Comparative Efficiency of Searching Titles, Abstracts, and Index Terms in a Free Text Data Base", Journal of Documentation, Vol. 28, No. 1, (March 1972), 13 pp. (See Document #3.)

Barker, Frances H., Douglas C. Veal, and Barry K. Wyatt. Retrieval Experiments Based on Chemical Abstracts Condensates, A Synopsis of the Final Report to OSTI on Project SI/21/47. (Nottingham, England: University of Nottingham, United Kingdom Chemical Information Service, July 1972), 13 pp.

Fisher, H. Leonard and Dennis R. Elchesen. "Effectiveness of Combining Title Words and Index Terms in Machine Retrieval Searches", Nature, Vol. 238, No. 5359, (July 14, 1972), pp. 109-110.

Garfield, Eugene. "Methods and Objectives in Judging the Information Content of Document Titles", Journal of Chemical Documentation, Vol. 10, No. 4, (November 1970), p. 260.

Hansen, Inge Berg. "CA Condensates as a Retrospective Search Tool: A Commentary", Information Storage and Retrieval, Vol. 9, No. 4, (April 1973), pp. 201-205.

Katzer, Jeffrey. "The Cost-Performance of an Online, Free Text Bibliographic Retrieval System", Information Storage and Retrieval, Vol. 9, No. 6, (June 1973), pp. 321-329. (See Document #24.)

Lancaster, F. Wilfrid, R. L. Rapport, and J. K. Penry. "Evaluating the Effectiveness of an Online Natural Language Retrieval System", Information Storage and Retrieval, Vol. 8, No. 5, (October 1972), pp. 223-245. (See Document #28.)

* * *

WYATT, Barry K. (See Documents #3, #52.)

* * *

53. YARBOROUGH, Judith. How to Prepare for a Computer Search of ERIC: A Non-Technical Approach. (Stanford, CA: Stanford University, ERIC Clearinghouse on Information Resources, September 1975), 44 pp. ED 110096.

The process of initiating and negotiating a literature search in the Education Resources Information Center (ERIC) data base is discussed in layman's terms. The guide offers suggestions on choosing a data base, deciding whether

to use computerized or manual retrieval, defining the search problem, and understanding the Boolean search strategy used in computer searches. It also presents a synthesis of the ERIC indexing guidelines that affect computer searches. Appendices include a list of data bases of interest to educators, a list of education level descriptors used in ERIC, a glossary, and a bibliography.

* * *